Application of Frailty-Based Mortality Models to Insurance Data

Zoltan Butt
and
Steven Haberman

Actuarial Research Paper No. 142

April 2002

ISBN 1 901615 65 0

Application of Frailty-Based Mortality Models to Insurance Data

Zoltan Butt
and
Steven Haberman

Actuarial Research Paper No. 142

April 2002

ISBN 1 901615 65 0

"Any opinions expressed in this paper are my/our own and not necessarily those of my/our employer or anyone else I/we have discussed them with. You must not copy this paper or quote it without my/our permission".

Application of Frailty–Based Mortality Models to Insurance Data

Zoltan Butt Steven Haberman

Department of Actuarial Science and Statistics
City University, London, U.K.

8 May 2001

Abstract

A range of frailty models — Makeham/Gompertz–gamma (with actuarial antecedents in the work of Perks and Beard), Gompertz–inverse Gaussian — has been considered and been applied to two insurance–based mortality data sets with the purpose of finding evidence of frailty in these populations and to determine, based on statistical tests, the consistency of such models for describing the age structure of mortality rates and recent trends over time. The models have been fitted using generalized linear models.

Keywords: *frailty, mortality rate, generalized linear modelling, gamma distribution, inverse Gaussian distribution, Gompertz–Makeham function, Perks model.*

1 Introduction

This paper considers practical aspects of the analysis of mortality in a heterogeneous population, and in particular focuses on selecting the frailty models that may be successfully fitted to insurance–based mortality data and on considering the implications of such models. The paper is organized in eight main sections.

The first section provides a general introduction to the topic of frailty models. The second section presents the framework used to formulate the mathematical models for representing frailty. The third section considers the selected models and different methods of fit. The fourth section considers goodness of fit and diagnostic tests. In sections five and six we present the insurance data and the results from fitting the range of models, their overall comparison and conclusions. In section seven, there is a brief discussion about the interpretation of parameter estimates of frailty–based mortality models and comparison of our results with other authors' findings. Section eight provides some concluding comments. Finally, a concise mathematical background is presented in the **Appendix.**

There is little doubt that heterogeneity exists in human population with respect to mortality. This is exemplified by studies of mortality in populations like the UK analyzed by socio–economic status, educational attainment, access to health system and other similar characteristics (see Benjamin and Pollard [1993]).

By the late 1970's, the theoretical implications of heterogeneity on population mortality studies became clearer to demographers, actuaries and statisticians working in this field. One particular feature was well established. As noted by Olshansky and Carnes [1997], both Gompertz and Makeham had recognized the limitations of the models for the age variation in mortality that they had proposed, and in particular the failure of these models to apply during the later part of the age span (see, for example, Makeham [1867] and Gompertz [1872]), which "has been a persistent theme throughout the historical literature on senescence and the search for a law of mortality".

Two distinct modelling approaches can be identified.

The first group of studies aimed at determining the effect of particular cause–specific factors on human mortality and thus assumed a mix of discrete terms when setting up mathematical models, for example, Levinson [1959], Redington [1969], Keyfitz and Littman [1979] and Congdon [1994].

The second group involved continuous rather than discrete risk models. Work on this approach effectively began with the seminal paper of Vaupel et al [1979] who formulated the basic ideas of quantifying the effect of heterogeneity on population mortality by the generalized concept of *frailty* represented as a *univariate* statistical variable $\boldsymbol{Z}$ and described its implications for standard life table methods.

The fundamental assumption is that each individual has a given vulnerability to death (or in fact to causes of death) and those more sensitive or ***frail*** will die sooner. Further, assuming that $\boldsymbol{Z}$ has a gamma distribution in the initial birth cohort across the population, Vaupel et al have shown that, as a result of heterogeneity, the mortality experience for an individual is significantly underestimated from population mortality trends by using previous methods, especially at older ages. The above concept can be readily extended to areas other than mortality, like disease incidence or component reliability, also to topics where *frail* does not necessarily mean inclined

to detrimental change, like migration or leaving unemployment (see, for example, Vaupel and Yashin [1985a]).

We propose a parametric model for the force of mortality using the frailty model as a building block. As a number of commentators have noted, there are many advantages in using parametric mortality models, including:

a) the resulting rates are smooth and can be used for interpolation or extrapolation by age;

b) construction of a complete survival model and associated life table is facilitated;

c) interpretation of parameters;

d) parsimony, so that extensive sets of rates or probabilities can be reduced to a few parameters (Keyfitz [1982]; Congdon [1993]; Chang [1998]);

e) analytic manipulation of functions based on the survival model;

f) comparison of mortality patterns between populations is facilitated;

g) assessment of trends and forecasting over time are assisted.

As is widely reported in the actuarial and demographic literature, the objective of using such parametric models is to obtain the best adherence to the data while retaining the smallest number of parameters. There is thus a trade off between goodness of fit and statistical stability of the parameters introduced (see, for example: Keyfitz [1982]; Benjamin and Pollard [1993] or Congdon [1993]).

In this paper, we will follow the method introduced by Vaupel et al [1979] and compare empirically two possible mortality models based on two different underlying frailty distributions, that is the gamma and inverse Gaussian, with the following aims:

a) to explore how substantial might be the heterogeneity present (via frailty) in different insurance–based populations;

b) to compare the behaviour of different mathematical models;

c) and to observe possible parameter trends.

2 Frailty Models

2.1 Model of Individual Differences in Frailty

Formulating a systematic way of considering the various aspects which distinguish individuals of a human population has proved to be difficult (Benjamin and Pollard [1993]). Even the current advances in computing and demographic registration,

greatly improving our data storage and recording expertise, solve just a small part of our problems. As noted by Keyfitz [1985], "...heterogeneity in the underlying population places difficulties in the way of interpretation of all statistical data based on averages". Practically, it is only feasible to form **partially homogeneous** subgroups since, if attempts are made to include all the factors in a descriptive or statistical model, the result would obviously be too over–parameterized and difficult to interpret. Thus, necessary simplifying assumption in traditional demographic (including life table) analysis has been to include only some of the important measurable differences in the composition of a population.

Standard methods of analysis tend to assume independence among individuals, while possible risk factors affecting mortality (or survival) may be covariates in the model. But there will be cases where we do not have complete, detailed information on each individual or where we do not know that the covariate exists, or if it exists, that it is of significance. These unobserved ***risk factors*** create dependence between the individuals under observation. In the statistical literature, models that cater for this dependence are often described as "random effects models".

Unlike the effect of sampling variability, which we can deal with by collecting more data, the effect of heterogeneity cannot be reduced by increasing sample sizes. An indirect approach based on the formulation of a model is needed to deal with heterogeneity as an omitted variable. The general consequence of ignoring heterogeneity is that the analysis of population data does not describe the characteristics of any individual but of the cohort as a whole. We note that, in mortality applications, the selective effect of heterogeneity is to underestimate the mortality rates at older ages which an average individual will experience. This bias will be explored further in this section.

Perhaps the first recognition of the effect of heterogeneity is due to Gini [1924] who was commenting on the interpretation of pregnancy rates. The effect in migration was considered by Blumen et al [1955] and in mortality by Vaupel et al [1979] which has led to considerable interest and further developments in the literature.

In the discussion that follows, we present the *'multiplicative model'* suggested by Vaupel et al [1979].

We define *frailty* in a heterogeneous population across any given age x as a random variable $\boldsymbol{Z}_x$ with probability density function $f_x(z)$, such that the individual force of mortality conditional on a realization of $\boldsymbol{Z}_x = z$ will satisfy the following relation:

$$\mu(x \,|\, \boldsymbol{Z}_x = z) \;=\; \mu_x(z) \;=\; z \cdot \mu_x \,, \tag{2.1}$$

where μ_x is referred to as the *'standard hazard function'* or *'standard force of mortality'* corresponding to a *'standard individual'*, conventionally those with frailty $z = 1$, that is: $\mu_x = \mu(x \,|\, \boldsymbol{Z} = 1)$. Observe that the meaning of μ_x is slightly changed from the usual actuarial notation. This model is presented in terms of the force of mortality, rather than the age specific probability of death, q_x, because of the fundamental

role played by the force of mortality in the survival model and because q_x is bounded above which would lead to complexities in terms of the range of values for $\boldsymbol{Z}_x$.

Therefore, we could consider *frailty* as an unobserved scale statistic (a continuous non–negative random variable) that encompasses <u>all</u> the factors affecting human mortality (other than age) in order to determine the extent of heterogeneity in a human population that is made up of individuals with different characteristics.

Further, assuming that there is a unique and fixed value of z associated with a given individual for all his or her life span, Vaupel et al [1979] note that "... The definition does not imply, however, that individuals at the same level of frailty are identical — even if they are contemporaries from the same population." The assumption is only that their "likelihood of death" would be the same. The result is that members of the cohort with higher values of frailty z will have a higher probability of dying and thus the more *frail* individuals are more likely to die first.

We also note that the assumption that frailty is constant for an individual during their lifetime is rather simplistic. However, this is the first step in incorporating heterogeneity in mortality analysis. Extensions to the model are briefly discussed in section 2.2[1].

We denote by H and S the cumulative hazard and the survivor function respectively for an individual up to a current age x. Then expressions for these functions, conditional on the frailty level z, are (by definition):

---

[1] In the above definition the 'standard' value conveniently and historically was chosen as 1. However, we note that the definition can be extended to any 'standard' value, say, $\boldsymbol{Z} = a$ by writing:

$$\mu_x{}' = \mu(x \,|\, \boldsymbol{Z} = a) = a \cdot \mu_x \,,$$

$$\therefore \mu(x \,|\, \boldsymbol{Z} = z) = z \cdot \mu_x = \frac{z}{a}\, \mu_x{}' = z' \cdot \mu_x{}' \,,$$

where $z' = z/a$ and $\mu_x{}'$ is the 'new standard' hazard function. Thus transforming the original definition (2.1) to:

$$\mu(x \,|\, \boldsymbol{Z} = az') = z' \cdot \mu_x{}' \,. \qquad (2.1')$$

Also in some models, we may have an extra constant term independent of age and frailty, for example, in the case of the Makeham individual hazard rate — see **Appendix**:

$$\mu_\alpha(x \,|\, \boldsymbol{Z} = z) = \alpha + z \cdot \mu_x \,. \qquad (2.1\alpha)$$

$$
\begin{aligned}
H(x\,|\,\boldsymbol{Z}=z) &= \int_0^x \mu(t\,|\,z)\,\mathrm{d}t \;=\; \int_0^x z\cdot\mu(t\,|\,1)\,\mathrm{d}t \\
&= \int_0^x z\cdot\mu_t\,\mathrm{d}t \;=\; z\int_0^x \mu_t\,\mathrm{d}t \\
&= z\cdot H_x\,, \qquad (2.2)\\
S(x\,|\,\boldsymbol{Z}=z) &= exp\,[-H(x\,|\,z)] \;=\; e^{-z\cdot H_x}\,. \qquad (2.3)
\end{aligned}
$$

where $H_x = \int_0^x \mu_t\,\mathrm{d}t$ is the 'standard' cumulative hazard function up to age x.

The probability density function (*p.d.f.*) of age at death among the survivors in the population at age x conditional on z follows from the definition of the force of mortality viz:

$$f(x\,|\,\boldsymbol{Z}=z) \;=\; f(x\,|\,z) \;=\; \mu(x\,|\,z)\cdot S(x\,|\,z) \qquad (2.4)$$

Therefore, the joint *p.d.f.* of the age at death variable and frailty z in the cohort is given by:

$$
\begin{aligned}
f(x,z) \;&=\; f(x\,|\,z)\cdot f_0(z) \;=\; \mu(x\,|\,z)\cdot S(x\,|\,z)\cdot f_0(z) \\
&=\; \mu(x\,|\,z)\cdot S(x,z)\,, \qquad (2.5)
\end{aligned}
$$

where we have used (2.4) and $S(x,z)$ is the individual survivor function dependent on both age x and frailty z:

$$S(x,z) \;=\; S(x\,|\,z)\cdot f_0(z)\,, \qquad (2.6)$$

and $f_0(z)$ is the *p.d.f.* of frailty in the initial birth cohort, that is at age $x=0$.

We can apply the above equations to the cohort at age x. For those in the population at age x, we note that death can occur at any time $t > x$, conditional on surviving until age x, and we can express the *p.d.f.* of variable $\boldsymbol{Z}_x$ (for those in the population alive at age x) as follows

$$f_x(z) \;=\; \frac{\int_x^\infty f(t,z)\,\mathrm{d}t}{\overline{S}(x)}\,,$$

where in the above

$$\overline{S}(x) \;=\; \int_0^\infty S(x\,,z)\,\mathrm{d}z \;=\; \int_0^\infty S(x\,|\,z)\cdot f_0(z)\,\mathrm{d}z$$

is the cohort survivor function **unconditional** on frailty, which can also be regarded as the proportion of the initial birth cohort still alive at age x regardless of their

frailty. Substituting using (2.1), (2.3), (2.5) and (2.6) gives:

$$\begin{aligned}\overline{S}(x)\cdot f_x(z) = \int_x^\infty f(t,z)\,\mathrm{d}t &= \int_x^\infty z\cdot\mu_t\cdot e^{-z\cdot H_t}\cdot f_0(z)\,\mathrm{d}t \\ &= f_0(z)\int_x^\infty z\cdot H_t'\cdot e^{-z\cdot H_t}\,\mathrm{d}t \\ &= f_0(z)\int_x^\infty \left[-e^{-z\cdot H_t}\right]'\mathrm{d}t \\ &= f_0(z)\cdot\left[-e^{-z\cdot H_t}\right]\Big|_x^\infty \\ &= f_0(z)\cdot e^{-z\cdot H_x} \\ &= f_0(z)\cdot S(x\,|\,z)\,, \end{aligned} \tag{2.7}$$

so that:

$$f_x(z) = \frac{f_0(z)\cdot S(x\,|\,z)}{\overline{S}(x)} = \frac{f_0(z)\cdot S(x\,|\,z)}{\int_0^\infty S(x\,|\,z)\cdot f_0(z)\,\mathrm{d}z}\,. \tag{2.8}$$

(We note that $\overline{S}(x)$ is the natural choice for the normalizing constant).
We denote the population hazard rate for individuals at age x by:

$$\overline{\mu}(x) = \frac{\int_0^\infty f(x,z)\,\mathrm{d}z}{\overline{S}(x)}\,. \tag{2.9}$$

Substituting from (2.5) and (2.6) yields:

$$\overline{\mu}(x) = \frac{\int_0^\infty \mu(x\,|\,z)\cdot S(x\,|\,z)\cdot f_0(z)\,\mathrm{d}z}{\overline{S}(x)}\,,$$

then substituting from (2.1) and (2.7) gives:

$$\begin{aligned}\overline{\mu}(x) &= \mu_x\int_0^\infty z\cdot f_x(z)\,\mathrm{d}z \\ &= \mu_x\cdot \mathrm{E}\left[\boldsymbol{Z}_x\right] = \mu_x\cdot\overline{z}_x\,, \end{aligned} \tag{2.10}$$

where $\mathrm{E}\left[\boldsymbol{Z}_x\right] = \overline{z}_x$ is the mean (or expected) frailty among the survivors to age x.

For the special case of (2.1α), we note that (2.10) becomes

$$\overline{\mu}(x) = \alpha + \mu_x\cdot\overline{z}_x\,. \tag{2.10α}$$

Choosing the marginal *p.d.f.* of z in the initial birth cohort such that its expected value is $\overline{z}_0 = 1$ (i.e. equal to the frailty level of the 'standard' individual [2]) will make

[2] Supposing that the 'standard' frailty is $z = a \neq 1$, then

substituting (2.1′) in: $\overline{\mu}(x) = \int_0^\infty \mu(x\,|\,z) f_x(z)\,\mathrm{d}z = \int_0^\infty {\mu_x}'\frac{z}{a} f_x(z)\,\mathrm{d}z = \frac{{\mu_x}'}{a}\overline{z}_x$

$\therefore$ if $\overline{z}_0 = a \Rightarrow \overline{\mu}(0) = \mu(0\,|\,z=a) = \mu_0'$; where μ_0' is the force of mortality of the 'standard' individual at birth.

the population hazard rate at birth equal to the 'standard' individual hazard rate:

$$\overline{\mu}(0) = \mu(0 \mid z = 1) = \mu_0 \,. \tag{2.11}$$

More importantly, we can see from the previous expression (2.10) that, should the expected frailty $\overline{z}_x$ vary significantly with age, the population force of mortality will have a different rate of change with respect to age than the individual force of mortality. Specifically, we note that:

$$\overline{z}_x = \frac{\int_0^\infty z \cdot f_0(z) \exp[-z \cdot H_x] \, \mathrm{d}z}{\int_0^\infty f_0(z) \exp[-z \cdot H_x] \, \mathrm{d}z} \,. \tag{2.12}$$

Differentiating (2.12) with respect to age x yields:

$$\frac{\mathrm{d}}{\mathrm{d}x} \overline{z}_x = -\mu_x \cdot \sigma_x^2(z) < 0 \tag{2.13}$$

where $\sigma_x^2(z)$ is the conditional variance of $\boldsymbol{Z}_x$ among the population that is alive at age x (Lancaster [1990], page 64). Thus, the mean frailty declines with age as death selectively removes those individuals with the higher frailty levels z who are likely to die earlier, and so $\overline{z}$ will decrease with age, especially in the case of the older ages. This means that μ_x increases more rapidly than $\overline{\mu}_x$ i.e. "individuals 'age' faster than heterogeneous cohorts" (Vaupel and Yashin [1985b], so that expectations of life for individuals would be overstated by not allowing for the selective effect of frailty. Regarding this feature Vaupel et al [1979] conclude: "... An intriguing implication is that studies of human aging based on cohort mortality data may be systematically biased or based on erroneous functional forms.".

However, in a microlevel analysis of mortality rates over space (borough) and time in London, Congdon [1994] finds no "universal tendency for a reduction in (life) expectancy when frailty is included." Congdon includes in his model covariates reflecting (inter alia) the ethnic composition of the sub-population and a deprivation index. Comparisons of life expectancy between a "degenerate" homogeneous model and a frailty–based model indicate the importance of differences in the parameter estimates for the underlying Gompertz model, the form of the heterogeneity model used and the specification of the level of over-dispersion in the variance. It would thus appear that, with a more complex model, the direction of the bias in measures of life expectancy is less clear-cut than with the simple frailty model presented above.

The model specified so far contains two critical components:

a) a model for describing the relationship between the hazard rate for individuals, μ_x, and current age x;

b) a choice for the distribution for $\boldsymbol{Z}_0$, $f_0(z)$.

The choices that have attracted the most attention in the literature are, for a), the Gompertz - Makeham model

$$\mu_x = \alpha + \beta \cdot e^{p \cdot x} \tag{2.14}$$

and for b) the gamma and inverse Gaussian distributions. Vaupel et al [1979] make extensive use of the gamma distribution on the grounds that it is tractable analytically, flexible in shape and has positive support (noting that frailty cannot be negative). The detailed implications of using these models are presented in the **Appendix.**

2.2 Extensions and Related Work

A number of contributions to the literature deal with adapting and fitting the model as presented in section 2.1.

Manton et al [1981] adapt equation (2.1) so that calendar time, y, is represented i.e.

$$\mu(x, y \,|\, \boldsymbol{Z} = z) = z \cdot \mu(x, y) \tag{2.15}$$

and then fit models with a range of choices for specifying $\mu(x, y)$ to national mortality data (grouped by age) for US and Sweden over the period 1850 – 1975. They conclude that the most satisfactory fit is obtained with the choice

$$\mu(x,\, y) = \mu(x) \cdot \exp\left[c_{y_0} + d_x\,(1885 - y_0)\right] \tag{2.16}$$

where c_{y_0} represents a contrast between the cohort born at time $y_0 = y - x$ and that born in 1885 and d_x is the change in the proportionality factor c for age x. In a further development, Vaupel [1999] surmises the following model for representing the dynamics of mortality over age and time:

$$\mu(x, y \,|\, Z = z) = z \cdot \mu(x) \cdot \exp\left[-r_1(x)\, r_2(y)\right] + \alpha \cdot \exp\left[-r_3(y)\right] \tag{2.17}$$

where $\alpha > 0$ and $r_1(x)$, $r_2(y)$ and $r_3(y)$ are suitably chosen functions.

Vaupel and Yashin [1985b] investigate the effects of different choices for $f_0(z)$ including a discrete model and uniform, gamma, Weibull and log–normal distributions.

Manton et al [1986] have explored the goodness of fit to US Medicare total mortality data and US lung cancer mortality data of a range of frailty models with the following choices for $f_0(z)$: gamma, inverse Gaussian and degenerate distributions (i.e. a homogeneous model) and for μ_x: Gompertz – Makeham and extended Weibull i.e.

$$\mu_x = \alpha + \beta \cdot x^{p-1} \text{ (for } \alpha \geq 0,\, \beta > 0,\, p > 0) \,. \tag{2.18}$$

They conclude that their results (in terms of parameter estimates and goodness of fit) are less sensitive to the choice of distribution for $f_0(z)$ than to the choice of function for μ_x.

Further levels of generality may be introduced by relaxing the assumption that the frailty level is constant with respect to time or age. Models allowing stochastic variation of frailty over time or age have been introduced and discussed by Woodbury and Manton [1977], Yashin et al [1985] and Vaupel et al [1988]. For example, Yashin et al [1985] differentiate between observed and unobserved variables (like frailty) — this is an important distinction in a life insurance context where, however rigorous the underwriting procedure, only some of the physiological variables for an individual can be measured and then *only* at policy inception.

The application to multiple state models, incorporating for example cause–specific forces of mortality, has been considered by Manton and Stallard [1980] and Vaupel and Yashin [1985b]. Also, Jones [1998] uses a four–state multiple state model with frailty for analyzing the impact of selective lapses on the mortality of insured lives.

A recent extension has been to formulate models of inherited frailty and longevity which can then be tested on mortality for related individuals (e.g. parent–child or twins): see Vaupel [1988], Yashin et al [1995] and Yashin and Iachine [1997].

3 Models and Methods of Fit

3.1 Introduction

We consider two methodologies for fitting the models described in the **Appendix**, the methods of non–linear least squares and of generalized linear modelling.

When selecting the mathematical models for graduating and fitting mortality data, it is important to consider not only achieving a reasonably good fit but also ensuring that the fundamental model assumptions are met. The classical method of non–linear least squares (**NLM**) requires the assumption of *constancy of variance* and the *normal distribution of error terms*, neither of which holds exactly in this case. Also, when modelling mortality or hazard rates the mean value should be strictly positive, whereas having a Gaussian distribution for the random responses implies, at least theoretically, a domain for the mean which is the whole real line. For these reasons, we use a generalized linear modelling (**GLM**) approach, which does not require these restrictive assumptions.

Further, it is well known that in the case of numbers of deaths when modelling mortality (as with data in the form of counts), the errors are expected to have a Poisson (and not normal) distribution. The Poisson distribution assumption is discussed more fully by Forfar et al [1988] and Renshaw [1991, 1995].

Therefore, only as a preliminary stage, graduation with respect to age by NLM (weighted and unweighted) has been compared to the different methods of GLM, which have been previously suggested by Renshaw [1991, 1995]. This testing has been carried out on a limited scale, being implemented only for the immediate annuitants' experience (see section 5.1). Then the finally selected GLM method has been applied

across the whole data set with a systematic sensitivity analysis, allowing for changes to the age range being fitted and the deletion of outliers. In the final phase, trends in the estimated parameters for the population hazard function, and indirectly for the chosen frailty distribution, have been investigated.

As noted earlier, we focus on the graduation of the crude forces of mortality by mathematical formula as this allows the definition of a frailty level as a continuous (and real) numerical value on the positive axis. The selected models under investigation are:

a) **Gompertz/Makeham – gamma:** $\overline{\mu}(x) = \alpha + \frac{a}{1+e^{b-p\cdot x'}}$, where $\alpha = 0$ in the case of the Gompertz individual hazard.

b) **Gompertz – inverse Gaussian:** $\overline{\mu}(x) = \frac{e^{-d+p\cdot x'}}{\sqrt{1+e^{-b+p\cdot x'}}}$.

In each model $x' = x - 40$. The full derivation of the above formulae can be found in the **Appendix**. We also observe that, later in this section, we use the simplifying notation of μ_x instead of $\overline{\mu}(x)$.

3.2 Exploratory Data Analysis

It is usual to undertake careful visual inspection of mortality rates plotted against age before attempting any graduation exercises. As noted by Horiuchi and Coale [1990] and others, it can be difficult to distinguish, from plots of the logarithms of mortality rates, the differences between a straight line and, say, upwardly or downwardly concave curve. These authors propose a different type of exploratory check. They consider the behaviour of

$$\lambda(x) = \frac{\mathrm{d}}{\mathrm{d}x} \ln \overline{\mu}(x) \tag{3.1}$$

for different curves. It is clear that if $\overline{\mu}(x) = \beta \cdot e^{p\cdot x}$ so that there is no heterogeneity present then $\lambda(x)$ would be exponentially increasing; however, for the Makeham – gamma frailty model

$$\overline{\mu}_\alpha(x) = \alpha + \frac{a}{1+e^{b-p\cdot x'}} \tag{A.13}$$

(see the **Appendix**) and then it is straightforward to demonstrate that if

$$\begin{cases} \alpha > 0, & \lambda(x) \text{ is bell shaped with a maximum at age } x_0 \\ \alpha = 0, & \lambda(x) \text{ is monotonically decreasing.} \end{cases}$$

(Indeed we can show that $x_0 - 40 = \frac{b}{p} + \frac{1}{2p} \ln\left(\frac{\alpha}{\alpha+a}\right)$).

For the Gompertz – inverse Gaussian model, the method can be similarly applied to:

$$\overline{\mu}(x) = \frac{e^{-d+p\cdot x'}}{\sqrt{1+e^{-b+p\cdot x'}}} \tag{A.19}$$

and it is straightforward to demonstrate that $\lambda(x)$ is monotonically decreasing.

Using some crude estimates for $\lambda(x)$, based on finite difference methods, for a sample of the data sets under consideration, we have noted that there is very little evidence of a bell–shaped profile, indicating the likelihood that $\alpha = 0$. The profiles also tend to be decreasing functions of age, indicating the presence of heterogeneity and supporting the choice of (A.13) with $\alpha = 0$ (i.e. (A.10)) or (A.19).

An equivalent test would be to consider a plot against age of

$$\lambda(x) = \ln\left[\overline{\mu}(x+1)^{-1} - \overline{\mu}(x)^{-1}\right] \tag{3.2}$$

which for the Gompertz – gamma model leads to a downward sloping straight line.

3.3 Non–Linear Least Squares

It is common to refer to a mathematical model as *non–linear* if there is at least one structural parameter that appears in all the partial derivatives (of first order and/or above) of the model with respect to any of its parameters. Such models generate parameter estimators of a more complicated form than linear models and usually can only be solved numerically. In this case we have made use of procedures offered by the S–PLUS 3.2 statistical package, see Venables and Ripley [1999], applying the non–linear least squares criterion, that is, minimizing the error sum of squares:

$$S(\boldsymbol{p}) = \sum_{i=1}^{n} w_{x_i} \cdot [\hat{\mu}_{x_i}(\boldsymbol{p}) - \mu_{x_i}]^2 \ , \tag{3.3}$$

with respect to the parameter vector $\boldsymbol{p}$, where the summation is over n cases (we use x_i as the notation for the observational units, which allows for the possibility that there might be grouped ages or some individual ages missing). The weights $w_{x_i} = 1$ for $\forall\, i = \overline{1,\, n}$ in the case of the unweighted model (NLM) and:

$$w_{x_i} = \frac{1}{\mu_{x_i}^2} \ , \ \forall\, i = \overline{1,\, n} \tag{3.4}$$

for the weighted graduation (WNLM). Hence the loss function becomes in this later case:

$$S(\boldsymbol{p}) = \sum_{i=1}^{n} \left[\frac{\hat{\mu}_{x_i}(\boldsymbol{p})}{\mu_{x_i}} - 1\right]^2 \ . \tag{3.5}$$

There is a considerable body of literature on the appropriate choice of weights in the non–linear least squares formulation: see Congdon [1993, 1994] for a discussion. One approach is to choose weights that are inversely proportional to the sampling variation ($\tau_{x_i}^2$, say) so that rates with greater variance (and lower reliability) attract lower weights. A second approach is to allow additionally for a constant level of overdispersion (σ^2) with particular reference to smaller populations, so that the

weights are proportional to $\left(\tau_{x_i}^2 + \sigma^2\right)^{-1}$. A third approach is to adopt (3.4), so that a constant coefficient of variation across all ages is assumed. Several investigators in the mortality field have followed this third approach, including Heligman and Pollard [1980], Forfar and Smith [1988], McNown and Rogers [1989], Carriere [1992] and Chang [1998].

We note that, with this configuration, cases with *zero* observed numbers of deaths, must be excluded from contributing to the loss function (3.5).

3.4 Generalized Linear Models

3.4.1 Introduction

The family of Generalized Linear Models (GLM) has been developed to apply a wide range of (usually) linear regression functions with an error structure which is different from the classical assumptions of a constant variance and a normal distribution. The underlying GLM principle is a natural extension of the classical linear model, with the independent variables present only through a linear configuration. The GLM method allows us to select the appropriate distribution of errors from a comprehensive class of density functions belonging to the exponential family and to define maximum likelihood estimators. The following brief introduction is based on McCullagh and Nelder [1989], while the actuarial applications of GLMs, (sections 3.4.3 and 3.4.4) have been inspired by Renshaw [1991, 1995] and Haberman and Renshaw [1996].

In the framework of GLM, the *generalization* of the classical linear regression models is carried out in two main directions:

1. **Linear Predictor Structure**

   The explanatory variables $\boldsymbol{x}_j$ $(j = \overline{1, p})$ control the mean of the response $\boldsymbol{m} = \mathrm{E}[\boldsymbol{Y}]$ through a linear combination $\boldsymbol{\eta} = \sum_j \boldsymbol{x}_j \beta_j$, termed the *linear predictor*, however the identity relationship $m_i = \eta_i$ is extended to any *monotonic* functional form:

   $$g(m_i) = \eta_i \qquad \text{or} \qquad m_i = g^{-1}(\eta_i) \qquad \forall\, i = \overline{1, n} \tag{3.6}$$

   where the function $g(\cdot)$ is called the *link function* being continuous and differentiable over its domain.

2. **Extended Distribution and Variance Assumptions**

   Similar to the traditional (linear or non–linear) regression analysis, an important feature of GLMs is to assume independently distributed random responses. However, the classical Gaussian distribution assumption is extended to a wider class of exponential density functions, also making the variance dependent on the mean of the response variable (i.e. non–constant) by a relation presumed to be known in advance.

Thus, GLMs assume that the response vector $\boldsymbol{y}$ is a random sample of the independent random variable $\boldsymbol{Y}$ with *p.d.f.*:

$$f_{\boldsymbol{Y}}(y,\, \theta,\, \varphi) \,=\, \exp\left\{\frac{y\cdot\theta - b(\theta)}{k(\varphi)} + c(y,\, \varphi)\right\}, \tag{3.7}$$

where $k(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are functions determined by the structure of the given GLM. For a known parameter φ, the above defines the exponential family with *canonical* parameter θ. The mean and variance of $\boldsymbol{Y}$ are then given by:

$$\boldsymbol{m} = \mathrm{E}[\boldsymbol{Y}] = b'(\theta) \qquad \text{and} \qquad \mathrm{Var}[\boldsymbol{Y}] = k(\varphi)\cdot b''(\theta)\,, \tag{3.8}$$

where we usually denote $b''(\theta) = V(m)$, also called the *variance function*, and observe that it depends on the mean $\boldsymbol{m}$ through the parameter θ.

The function $k(\cdot)$, is a scale factor applied to the variance and usually is expressed as a ratio:

$$k_i(\varphi) = \frac{\varphi}{\omega_i} \qquad \forall\, i = \overline{1,\, n} \tag{3.9}$$

that is a ratio between a constant *dispersion* parameter φ and some predefined w_i weights, potentially different for each observation.

It is interesting to note that we can obtain an optimal GLM for the special link function referred to as the *canonical link* defined by the condition $g(m_i) = \theta_i$. Examples are the identity, log and log-odds link functions for the normal, Poisson and binomial distributions respectively. Then, as one would expect, the classical linear modelling structure acts as a special (limiting) case of the GLM regression analysis having *identity* link $g(\boldsymbol{m}) = \boldsymbol{m}$ and variance function $V(\boldsymbol{m}) = 1$.

To fit parameterized models represented by $\boldsymbol{\eta}$, the unknown parameters are estimated by maximizing the log–likelihood. A measure of the overall goodness of fit of the current model under consideration is provided by the value of the scaled deviance, which is defined to be twice the difference between the log–likelihood achieved by the *saturated* (i.e. matches the observations exactly) and the *fitted* model:

$$\begin{aligned} D^*(\boldsymbol{y};\hat{\boldsymbol{m}}) \,&=\, 2\left\{l(\boldsymbol{y},\, \tilde{\boldsymbol{\theta}},\, \varphi) - l(\boldsymbol{y},\, \hat{\boldsymbol{\theta}},\, \varphi)\right\} \\ &=\, 2\left\{\sum_{i=1}^{n}\left[\frac{y_i\cdot\tilde{\theta}_i - b(\tilde{\theta}_i)}{k_i(\varphi)} + c(y_i,\, \varphi)\right] - \sum_{i=1}^{n}\left[\frac{y_i\cdot\hat{\theta}_i - b(\hat{\theta}_i)}{k_i(\varphi)} + c(y_i,\, \varphi)\right]\right\} \\ &=\, \frac{1}{\varphi}\sum_{i=1}^{n} 2\,\omega_i\left[y_i\cdot(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)\right]\,. \end{aligned} \tag{3.10}$$

In the above, the canonical parameter θ corresponding to the fitted and saturated (or full) model is denoted by $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}(\hat{\boldsymbol{m}})$ and $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}(\boldsymbol{y})$ respectively. The saturated model is characterized by the property that it provides a perfect fit, since the fitted values are equal to the empirical responses, themselves. In the GLM method, it is

customary to work with the unscaled deviance $D(\cdot)$ instead of the scaled one $D^*(\cdot)$, which is defined as:

$$D(\boldsymbol{y};\, \hat{\boldsymbol{m}}) \;=\; \varphi \cdot D^*(\boldsymbol{y};\, \hat{\boldsymbol{m}})\,. \tag{3.11}$$

We note that the scaled deviance is a measure of discrepancy equivalent to that used in the classical regression analysis (based on the normal error distribution), namely the error sum of squares and also, in this case, the scale factor is equal to the variance, that is $\varphi = \sigma^2$.

Differences in the values of the scaled deviance, as more terms are added to the parameterized structure of the predictor, are then used to assess the statistical significance of the addition to the structure. Such deviance differences may be referred, as an approximation, to the χ^2 distribution with the appropriate degrees of freedom determined by the number of extra parameters which have been added to the nested structure of the predictor $\boldsymbol{\eta}$.

The scale parameter φ is estimated either by the ratio

$$\hat{\varphi} \;=\; \frac{D(\boldsymbol{y};\hat{\boldsymbol{m}})}{\nu} \tag{3.12}$$

based on the unscaled deviance, measure of goodness of fit, or by the ratio

$$\hat{\varphi} \;=\; \frac{1}{\nu}\sum_{i=1}^{n} \omega_i \frac{(y_i - \hat{m}_i)^2}{V(\hat{m}_i)}\,. \tag{3.13}$$

based on the unscaled Pearson goodness of fit statistic. Here ν denotes the degrees of freedom associated with the model fit.

Apart from the usual *response* residuals $\boldsymbol{r} = \boldsymbol{y} - \hat{\boldsymbol{m}}$ there are three main types of residuals available in the GLM framework, referred to as *deviance*, *Pearson* and *Anscombe* residuals. In the current paper we focus only on the first two, each of these being in effect the components of the corresponding *unscaled* discrepancy measures:

i. **Deviance Residuals**

In the formal definition of the deviance (3.10) and (3.11) we can see that $D(\boldsymbol{y};\, \hat{\boldsymbol{m}}) = \sum d_i$, so by defining:

$$\boldsymbol{r}_D \;=\; \operatorname{sign}(\boldsymbol{y} - \hat{\boldsymbol{m}})\sqrt{\boldsymbol{d}}\,, \tag{3.14}$$

we will have $\sum r_D^2 = D(\cdot)$.

ii. **Pearson Residuals**

Similarly to before, we define this as a quantity that satisfies $\sum r_P^2 = X^2$, based on (3.13):

$$\boldsymbol{r}_P \;=\; (\boldsymbol{y} - \hat{\boldsymbol{m}})\sqrt{\frac{\boldsymbol{\omega}}{V(\hat{\boldsymbol{m}})}}\,. \tag{3.15}$$

3.4.2 Modelling Mortality with Poisson Distribution

In the traditional actuarial approach, the force of mortality is assumed to be constant over the age interval under consideration, usually 1 year, and further the deaths of members of the cohort under observation are assumed to be independent.

The observed data comprise the number of recorded deaths, a_x, arising from central exposed to risk, r_x, over a range of ages x in a defined calendar period. As noted earlier, our target is to model the force of mortality at age x, μ_x. Following Forfar et al [1988], we model the actual numbers of deaths, A_x, as Poisson random variables with mean equal to $r_x \cdot \mu_x$. Forfar et al [1988] show that the maximum likelihood estimator of the force of mortality is:

$$\hat{\mu}_x = \frac{A_x}{r_x} \,. \tag{3.16}$$

With a one year age interval $(x,\ x+1)$, it has become customary to refer the estimator to the central age of the interval i.e. $\hat{\mu}_{x+\frac{1}{2}}$. In the following paragraphs, we use the suffix x_i to refer to the i^{th} age group, thereby retaining generality, and we overlook the need to refer to the central age of the interval in places for notational convenience.

Therefore, our aim is to model mortality, making use of the GLM structure based on the Poisson distribution of the response variable $\boldsymbol{Y} \sim \mathrm{P}(\boldsymbol{m})$, the numbers of death in the age interval under discussion, with *p.d.f.* of the form:

$$f(y,\, \theta,\, \varphi) \;=\; exp\,\{y \cdot \log m - m - y!\} \;. \tag{3.17}$$

A particular feature of the data to be used is that they are based on the number of policies rather than the number of lives. So the death of a policyholder with more than one policy will appear in the observations as more than one death. This feature needs to be allowed for in the model. It has received wide attention in the literature: see Forfar et al [1988] for example. We follow the proposal of Renshaw [1992], and adopt a modelling distribution that incorporates overdispersion i.e. the overdispersed Poisson distribution. We note, however, that the modelling of excess variation (as in this case) has only a marginal effect on the parameter estimates (as we might intuitively expect) but that confidence intervals and significant tests may be biased unless the effect is incorporated: see Cox [1983] and Renshaw [1992] for further discussion.

Then, the statistics of the corresponding GLM are given by:

$$\left.\begin{array}{c} \mathrm{E}[Y_i] \;=\; m_i \;=\; r_{x_i} \cdot \mu_{x_i} \qquad \text{and} \qquad \mathrm{Var}[Y_i] \;=\; k_i(\varphi) \cdot m_i \\ \theta_i \;=\; \log m_i \qquad b(\theta_i) \;=\; \exp\theta_i \qquad V(m_i) \;=\; m_i \\ \text{canonical link:} \qquad g(m_i) = \theta_i = \log m_i \\ r_{D\,i} \;=\; \mathrm{sign}\,(y_i - \hat{m}_i)\,\sqrt{2\left[y_i \ln \frac{y_i}{\hat{m}_i} - (y_i - \hat{m}_i)\right]} \\ r_{P\,i} \;=\; (y_i - \hat{m}_i)\sqrt{\frac{1}{\hat{m}_i}} \end{array}\right\} \tag{3.18}$$

where $k_i(\varphi) = \varphi$.

In this paper, we have applied the GLM fitting method to the adopted Gompertz – gamma frailty model. The method of implementation depends on the chosen target variable, which could be either the number of deaths or the force of mortality, so that $Y_i = A_{x_i}$ or $Y_i = A_{x_i}/r_{x_i}$ respectively, and these two methods are outlined in the following two sections. In both, we have made the simplifying assumption that the central exposures are constant (non–random) quantities for each observational unit.

The actual fitting applied is based on the so–called 'quasi–log–likelihood' technique, that defines the quantity:

$$Q(\boldsymbol{m};\, \boldsymbol{y}) \;=\; \sum_{i=1}^{n} q(m_i;\, y_i) \;=\; \sum_{i=1}^{n} \omega_i \int_{y_i}^{m_i} \frac{y_i - t}{\varphi\, V(t)}\, \mathrm{d}t \tag{3.19}$$

where y_i are the observations with mean m_i. Therefore the corresponding (quasi) unscaled deviance is defined as:

$$D_q(\boldsymbol{y};\, \hat{\boldsymbol{m}}) \;=\; -2 \sum_{i=1}^{n} \varphi \cdot q(\hat{m}_i;\, y_i) \;=\; -2 \sum_{i=1}^{n} \omega_i \int_{\hat{m}_i}^{y_i} \frac{y_i - t}{V(t)}\, \mathrm{d}t\,. \tag{3.20}$$

This method represents an extension of the fundamental approach presented earlier in section 3.4.1, since, for members of the exponential family of distributions, the quasi–log–likelihood can be shown to be equivalent to the log–likelihood. McCullagh and Nelder [1989] provide a detailed mathematical treatment of quasi–log–likelihood functions.

3.4.3 GLM with Taylor Expansion

The modelling structure described so far (which we shall call GLM1) considers the number of deaths $\boldsymbol{A}$ as the target random variable having a Poisson distribution for each individual case, so that:

$$A_x \sim \mathrm{P}\left(r_x \cdot \mu_{x+\frac{1}{2}}\right) \;\Rightarrow\; m_x \;=\; \mathrm{E}\left[A_x\right] = r_x \cdot \mu_{x+\frac{1}{2}}\,, \tag{3.21}$$

where the expected value is denoted by m_x as above.

The corresponding GLM features result by substituting the given mean m_x into equations (3.18). However, note that we have included in this model an overdispersion parameter $\varphi > 1$, to allow for the possible presence of duplicate policies. Thus, the variance is:

$$\mathrm{Var}\left[A_x\right] = \varphi \cdot V(m_x) = \varphi \cdot r_x \cdot \mu_{x+\frac{1}{2}}\,. \tag{3.22}$$

As mentioned before the final model of $\overline{\mu}(x)$ is non–linear in some of its parameters, so the above relation (3.21) for the mean number of deaths m_x has to

be *'linearized'* in order to be able to apply the GLM methodology. We follow Renshaw [1991] and first take the logarithmic transformation of (3.21) and then use a Taylor expansion for the remaining non–linear term in order to obtain an approximation to the mean, which is linear in all of the resulting structural parameters. Thus,

$$\begin{aligned} \ln m_x &= \ln r_x + \ln \mu_{x+\frac{1}{2}} \\ &= \ln r_x + \ln a - \ln\left(1 + e^{b - p\cdot x'}\right) , \end{aligned} \tag{3.23}$$

where $x' = (x + \frac{1}{2} - 40)$. We denote by $f(b, p)$ the last term in the above and we apply a Taylor expansion, to first order terms, in the neighbourhood of (b_0, p_0) to obtain

$$\ln m_x \simeq \underbrace{\ln r_x - c_x}_{\text{offset}} + \ln a + \beta \cdot k_x + \gamma \cdot v_x , \tag{3.24}$$

where $c_x = \ln(1 + e^{b_0 - p_0 \cdot x'})$, $k_x = e^{b_0 - p_0 \cdot x'}(1 + e^{b_0 - p_0 \cdot x'})^{-1}$, $v_x = x' \cdot k_x$, $\beta = b_0 - b$ and $\gamma = p - p_0$. The RHS of (3.24) then forms a linear predictor with the unknown parameters $\ln a$, β and γ with a known offset $\ln r_x - c_x$.

The estimates from this GLM procedure depend on a set of starting values for b_0 and p_0 so that c_x, k_x and v_x can be computed. The closer these values are to the global minimum, the more certain are we that the iteration will converge to the right set of estimated values. Since, in the present case, we cannot estimate b and p directly, the actual fitting is performed recursively by reusing the estimates of $\hat{b}$ and $\hat{p}$ as starting values for the next fitting:

$$\hat{b} = b_0 - \hat{\beta} \qquad \text{and} \qquad \hat{p} = p_0 + \hat{\gamma}$$

until $\hat{\beta} \simeq 0$ and $\hat{\gamma} \simeq 0$, that is the estimates of $\hat{b}$ and $\hat{p}$ cannot be improved upon. We use, as starting conditions, estimates from previously fitted observation periods or the results from other graduation methods (for example, NLM or WNLM).

3.4.4 GLM with Parameterized Link Function

In the second GLM graduation method (GLM2) the force of mortality is selected as the target random variable, and we use the overdispersed Poisson distribution to model the numbers of death, as in the previous sections. Considering the responses to be A_x/r_x, the mean and variance can be expressed as:

$$\mathrm{E}\left[\frac{A_x}{r_x}\right] = \frac{\mathrm{E}[A_x]}{r_x} = \frac{m_x}{r_x} = u_x , \tag{3.25}$$

$$\mathrm{Var}\left[\frac{A_x}{r_x}\right] = \frac{\mathrm{Var}[A_x]}{r_x^2} = \frac{\varphi \cdot m_x}{r_x^2} = \frac{\varphi}{r_x} u_x \tag{3.26}$$

where the notation u_x has been used to highlight the difference between the mean of the two target variables, $u_x \neq m_x$. Comparing the above with the configuration of

the GLM (3.8) and (3.9) we can see that in order to satisfy the Poisson distribution assumptions we have to set the weights as follows:

$$\omega_x = r_x \, .$$

Then the Gompertz – gamma form of the frailty model (A.10) can be rewritten in terms of the mean value as:

$$u_x = \frac{a}{1 + e^{\eta_x}} \, , \tag{3.27}$$

where $\eta_x = b - p \cdot x'$. Although it is not possible to form a linear predictor η_x that would include all three parameters a, b and p, we can apply the GLM methodology, if we are able to find a continuous and differentiable link function $g(u_x, a) = \eta_x$, with fixed parameter a. Then, the graduation could be performed over a chosen range of values for parameter a assuming that the resulting *'deviance profile'* $D(\boldsymbol{\mu}; \hat{\boldsymbol{u}}_a)$ is a smooth curve with preferably one (global) minimum value $D_{\min}$ over the selected interval.

Observe that we can modify the original expression (3.27) by isolating $\exp(\eta_x)$ and then taking logarithms of both sides:

$$\ln\left(\frac{a - u_x}{u_x}\right) = \eta_x \, ,$$

resulting in a new *'parameterized link'* function:

$$g(u_x, a) = \ln\left(\frac{a - u_x}{u_x}\right) \, . \tag{3.28}$$

This approach has been used by Renshaw [1995]. The value of the parameter a has to be pre–determined before the GLM can be fitted. For the structure under consideration, it is possible to search for the optimum value of a by refitting the same predictor structure η_x for different values of a, chosen carefully so that a deviance profile can be constructed.

Note that there is a lower bound imposed on the values which the parameter a can take, since the argument of the logarithm must be positive, that is $a > \max(u_x)$ (see also the interpretation of a in section 7.2). Although we do not attempt to prove here that the resulting deviance curve $D(\boldsymbol{\mu}; \hat{\boldsymbol{u}}_a)$ is indeed continuous and has a single minimum over the domain $(a_{\min}, \infty]$, the results from extensive trials on a range of data sets indicate that this is the case. Figure 1 provides an example of deviance profiles based on the Female Annuitants' Ultimate (i.e. policy duration over 5 years) experience for a range of 8 calendar periods.

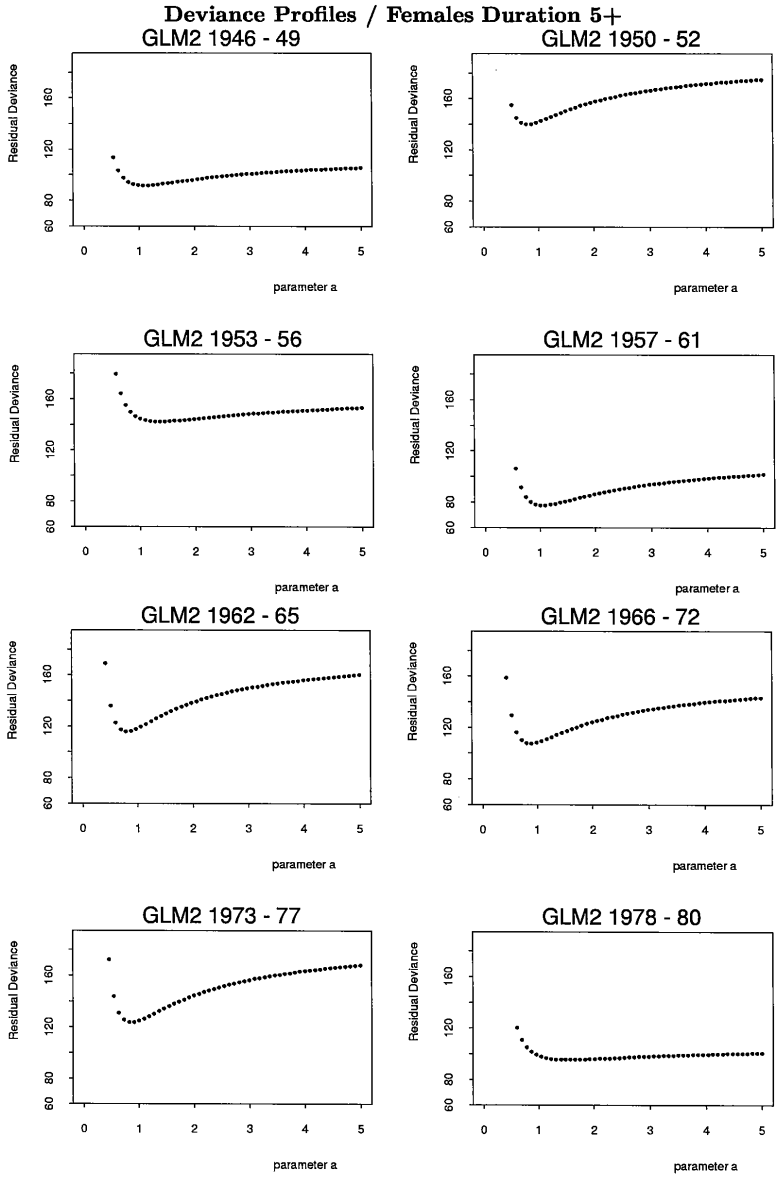

Figure 1: Deviance vs. fixed values of parameter a for Female Annuitants.

4 Testing of Goodness of Fit

Appropriate diagnostic checks have been applied to all the graduated models, based both on a visual inspection of characteristic plots and computation of the test statistics proposed and implemented in the actuarial field by the CMI Bureau and codified by Forfar et al [1988] (i.e. signs test, runs test, serial correlation test). A typical graphical diagnostic analysis involves plots of residuals against fitted values and against observed values, plots of the histograms of residuals and also Normal QQ plots (as proposed by Renshaw [1991, 1995]). All the statistical tests are based on the Pearson residuals r_P.

5 Application to Insurance Data

The proposed models have been applied to two major sets of insurance data collected by the CMI Bureau and available for an extended historical period. The data comprise tabulations of observed number of deaths a_x and initial exposures r_x^i by age x and policy duration d. Since we have modelled forces of mortality (μ_x) instead of probability of deaths (q_x) the original initial exposure counts have been transformed to central exposures by the (standard) approximate formula:

$$r_x = r_x^i - \frac{1}{2} a_x \,. \tag{5.1}$$

5.1 Immediate Annuitants

The data consist of cohorts of males and females with lives born in the period between 1846–1943 and followed up between 1946–1994 (note that the observations for years 1968, 1971 and 1975 are not available). The data have been originally recorded by grouped ages 40–45, 45–50, 100+ and for individual ages between 50 – 100, and also by duration periods (1–4 years and 5+, subsequently rearranged as 1+, i.e. 1–4 and 5+, i.e. 5 and over). Also, further grouping by observation years (and some ages) has been applied in order to reduce the occurrence of zero mortality rates in individual cells; this is particularly important for the application of the *weighted non linear least squares* graduation method (see section 3.3).

We have grouped the observation years so that the central year of any group does not coincide with the end of a decade (to avoid bias):

1946 – 49 (4)	**1950 – 52 (3)**	**1953 – 56 (4)**	**1957 – 61 (5)**
1962 – 65 (4)	**1966 – 72 (7-2=5)**	**1973 – 77 (5-1=4)**	**1978 – 80 (3)**
1981 – 85 (5)	**1986 – 90 (5)**	**1991 – 94 (4)**	

where the numbers in parentheses are the number of participating observation years. Graduation has been carried out principally on individual ages between 50 and 100+

with minor changes from one method to another. We note that, for immediate annuitants, the female data set is larger than the male data set and we focus here on results for the female data.

5.2 Assured Lives (Males)

This was the largest data set analyzed and it was in a similar format to the annuitants' data set. The experience of male policyholders, however, dominates and is the focus of our analysis. The recorded age range is considerably longer than the annuitants' data set — generally individual ages between 10 – 100+, although models have only been applied to adult ages 50 – 100+. The available duration periods have been used in the form 0, 1, 2–4 and 5+, in line with other authors (for example, Renshaw and Haberman [1997]). We note that the data in the first three groups, corresponding to lower durations, are significantly smaller in size and less reliable than the 5+ group. The data are available for groups of 4 years, with the exception of the last 4 calendar years 1991– 94, where the data for individual calendar years are also given:

1924 – 28 (5)	**1929 – 33 (5)**	**1934 – 38 (5)**	**1949 – 52 (4)**
1953 – 58 (6)	**1959 – 62 (4)**	**1963 – 66 (4)**	**1967 – 70 (4)**
1971 – 74 (4)	**1975 – 78 (4)**	**1979 – 82 (4)**	**1983 – 86 (4)**
1987 – 90 (4)	**1991 – 94 (4)**		

As before, the numbers in parentheses are the number of contributing years.

6 Results and Parameter Estimates

In this section, we illustrate the graduation results for the models selected. Given the similarity between some of the results and the different possible criteria for consideration when filtering the final model, it should be stressed that there is often not a single model which provides an optimal fit from all points of view. Full details of the results can be obtained from the authors.

6.1 Annuitants Experience

In the first stage, all the models presented have been fitted to the annuitants' experience using all the available data points with both graduating techniques (NLM and GLM), and consideration has been given not only to the final parameter estimates and to the goodness of fit tests, but also to the sensitivity of the results to changes in the starting conditions. We note that in some cases when graduating by the weighted classical non–linear least squares method (WNLM), further grouping by ages of the data cells is necessary in order to avoid *zero* observed deaths and thus *infinite* weights.

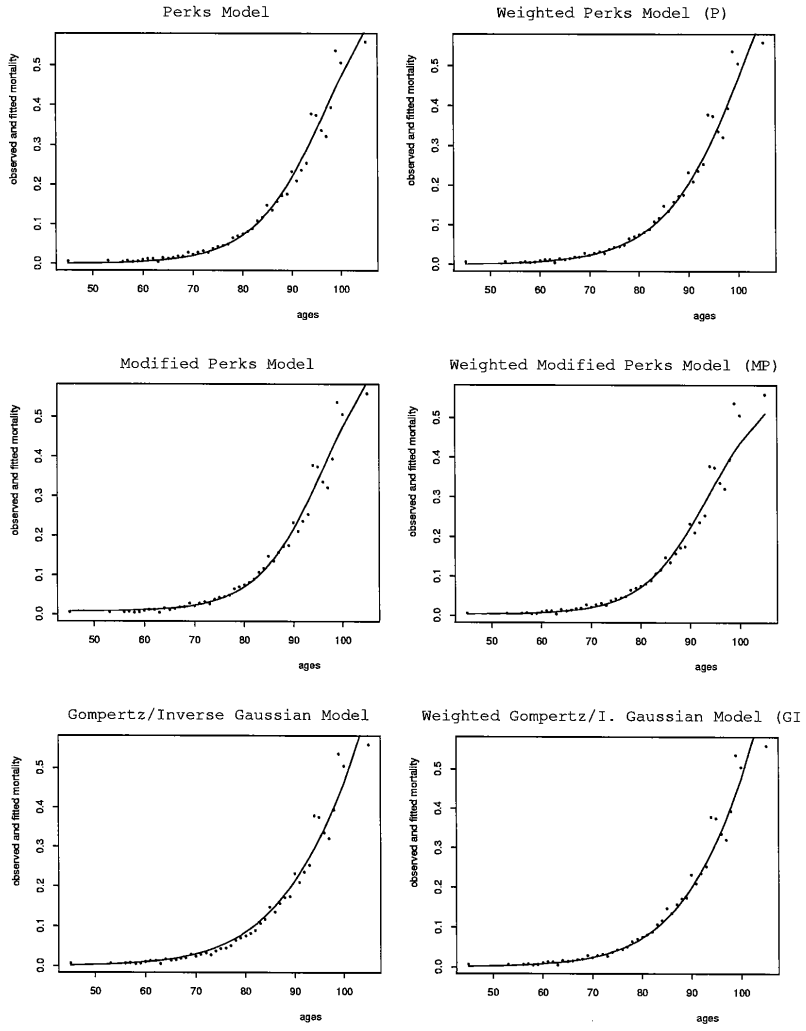

Figure 2: Predicted and observed forces of mortality for Female Annuitants Duration 5+ 1953 – 56. Perks, Modified Perks and Gompertz-inverse Gaussian models fitted by unweighted and weighted least squares.

As noted earlier, we focus on the female immediate annuitants' experience. The models considered are:

a) **Gompertz – gamma, known as Perks (P)**

$$\overline{\mu}(x) = \frac{a}{1 + e^{b - p \cdot x'}} ; \tag{A.10}$$

b) **Gompertz/Makeham – gamma, known as Modified Perks (MP)**

$$\overline{\mu}(x) = \alpha + \frac{a}{1 + e^{b - p \cdot x'}} ; \tag{A.13}$$

c) **Gompertz – inverse Gaussian (GI)**

$$\overline{\mu}(x) = \frac{e^{-d + p \cdot x'}}{\sqrt{1 + e^{-b + p \cdot x'}}} . \tag{A.19}$$

As an example of the results, we present Figure 2 which shows predicted and observed forces of mortality for the female annuitants' experience for 1953 – 56 for duration 5+ with fitting carried out by the unweighted (NLM) and weighted non–linear least squares methods (WNLM) of section 3.3. Similar graphs (not shown here) are available for the other subdivisions of the data set.

Comparing initially the results for the WNLM regressions, it is concluded that it would not be possible to distinguish between the models based on the goodness of fit tests alone. However, from the detailed parameter estimates and their standard errors, we find that the $\hat{\alpha}$ estimates of the MP model are not significantly different from zero and in some cases are negative. Similarly, the $\hat{a}$ parameter estimates for the P and MP models are, in some cases, not significant. On the other hand, all of the parameter estimates for the GI model are significant. Therefore, based only on this criterion one might select the GI as the best model for the given data set, especially when considering the 'favourable' property of a non–constant coefficient of variation for the inverse Gaussian (see equation (A.17) and related comments in the **Appendix**). However, it is generally accepted that the population hazard is best described by a sigmoidal shape curve (i.e. logistic family function), which for older ages tends to a positive finite value (Thatcher [1999]). The GI model does not have this characteristic and it reduces to an exponential function, i.e. from the Gompertz family, as $x \to \infty$: for example, see the fitted models to the grouped observation years 1953 – 56 in Figure 2 for females and Figure 7 (later) for males.

Examples of the diagnostic plots are presented in Figures 3 and 4 for the P and GI model fitted to the 1950 – 52 and 1991 – 94 data for duration 1+. The plot of the residuals against age and fitted values (in the first 2 panels) should show an absence of any systematic pattern. Further, based on the assumption of a normal error distribution, the QQ Normal plot of the residuals should be a straight line. These plots indicate a poor fit to the data for ages in excess of 85, in this case,

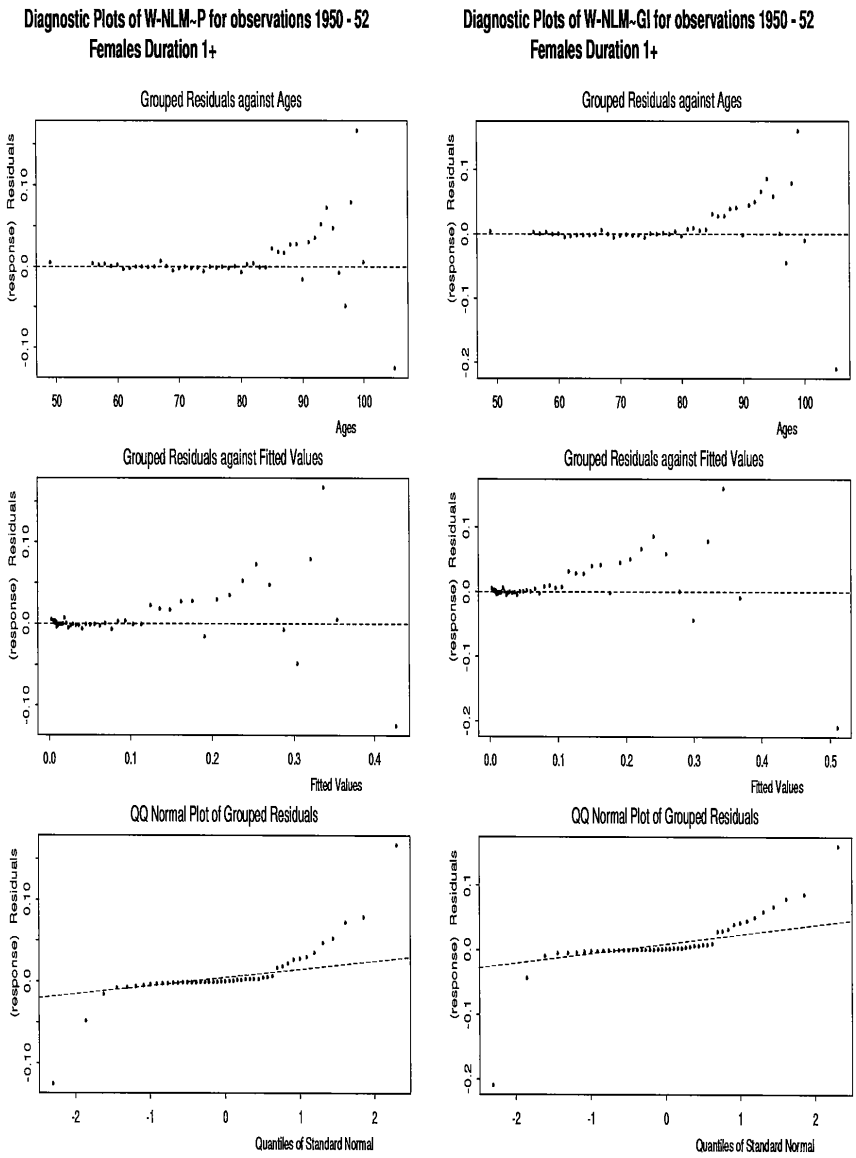

Figure 3: Comparison of WNLM fitting of Perks and Gompertz – inverse Gaussian models for Female Annuitants Duration 1+ 1950 – 52.

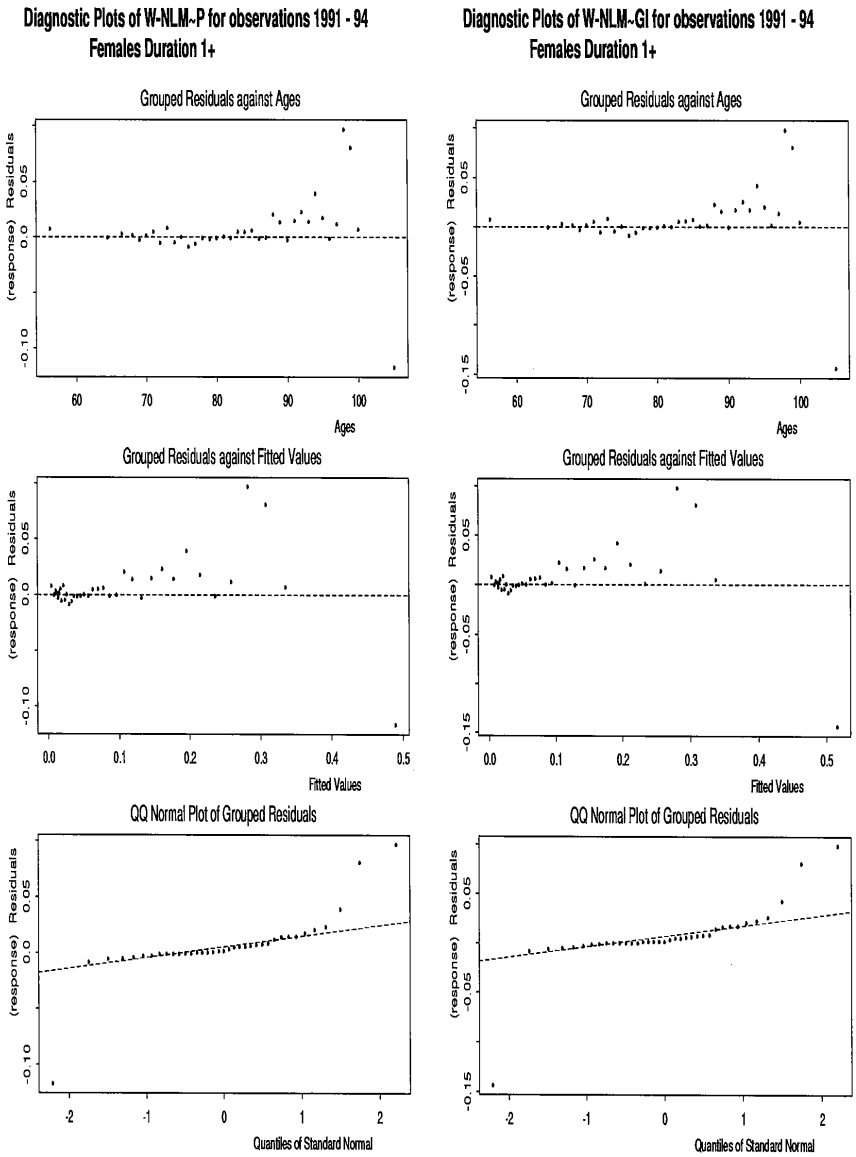

Figure 4: Comparison of WNLM fitting of Perks and Gompertz – inverse Gaussian models for Female Annuitants Duration 1+ 1991 – 94.

for both models. The results are even more striking when considering the shape of the *unweighted* regressions for some of the more *'sparse'* data sets like the male annuitants (Figure 7) illustrating further the rigidity of the GI model compared to the P or MP models, with the GI model generally having poorer goodness of fit properties in these cases. In general, the P model produces a better fit to the data than the GI model. Therefore, we have opted, on balance, for the P model from the three initially proposed population hazard functions (A.10), (A.13) and (A.19), while keeping in mind that our particular aim is then to use these results as a starting point for taking advantage of the improved regression properties of the GLM method.

Although both GLM–based methods presented in sections 3.4.3 and 3.4.4 provide similar results we only focus here on the GLM2 method, this being the more straightforward for estimating parameter standard errors. We can obtain a good general impression of the features of the fitted models by looking at the estimated parameters of the individual hazard functions and the assumed frailty distributions.

We note from the **Appendix** that the Gompertz–gamma (or Perks) model is of the form of (A.9) where $\delta = \theta$ is the shape parameter for the gamma distribution, with the mean equal to 1, that is:

$$\overline{\mu}(x) = \frac{\delta \cdot p \cdot \beta \cdot e^{p \cdot x}}{(\delta \cdot p - \beta) + \beta \cdot e^{p \cdot x}},$$

where the Gompertz model $\mu_x = \beta e^{p \cdot x}$ is used for the standard individual.

The connection between these fundamental parameters and the form (A.10) used in the fitting is:

$$\hat{\beta} = \frac{\hat{a}}{1 + e^{\hat{b} + 40\hat{p}}}, \tag{A.11}$$

$$\hat{\delta} = \hat{\theta} = \frac{\hat{a}}{\hat{p}}, \tag{A.12}$$

The parameter p is common to both parameterizations.

Similarly, for the Gompertz–inverse Gaussian case, the model is of the following form, with $\gamma = \psi$ so that the mean of the inverse Gaussian distribution is equal to 1:

$$\overline{\mu}(x) = \beta \cdot e^{p \cdot x} \left(\frac{\gamma}{\gamma + \frac{\beta}{p}(e^{p \cdot x} - 1)} \right)^{\frac{1}{2}}.$$

The connection between these fundamental parameters, p, β and γ and the form (A.19) used in the fitting is:

$$\hat{\beta} = \frac{e^{-\hat{a} - 40\hat{p}}}{\sqrt{1 + e^{-\hat{b} - 40\hat{p}}}}, \tag{A.20}$$

$$\hat{\gamma} = \hat{\psi} = \frac{e^{-\hat{a} - 40\hat{p}}}{e^{-\hat{b} - 40\hat{p}}} \frac{\sqrt{1 + e^{-\hat{b} - 40\hat{p}}}}{\hat{p}}, \tag{A.21}$$

The parameter p is common to both parameterizations.

We present these values based on the WNLM and GLM2 approaches for the female annuitants data set with two duration periods (1+ and 5+) and 11 grouped observation periods in Figures 5 and 6. We can see that the estimates $\hat{\beta}$ and $\hat{p}$ of the GI and P models have an almost identical pattern over time in Figure 5, confirming the similarity of these two models for the given age range (as previously suggested based on the goodness of fit tests), whereas the MP model differs significantly from these. It is interesting to note that the estimates based on the GLM2 method for the P model are within the same range as those given by the GI and P models fitted by the NLM method, but display a much smoother variation by period. Also, in the case of the frailty distribution parameter ($\hat{\theta}$ or $\hat{\gamma}$), all the models lead to the same type of pattern but at different levels. In the case of Figure 6, corresponding to durations 5+, the GLM2-P model results in more stable estimates.

As described earlier in section 3.1, in the case of mortality modelling, we would prefer a model that satisfies reasonably well (among others) two important criteria, namely:

a) allows for the inhomogeneity in the variance of the errors;

b) has parameter estimates that are not influenced by the outliers in the data.

Then, the better behaviour of the GLM2 regression method could be associated with the fact that it is less influenced (than the WNLM method) by the outliers of a given data set as demonstrated by the results for 1953 – 56 males' experience in Figures 7 and 8, which show that the P and MP models are strongly influenced by the presence of outliers in the WNLM case (Figure 7) and that the P model is less influenced by the presence of outliers in the GLM2 case (Figure 8). Although the NLM method is as equally unaffected by the outliers as the GLM2, i.e. behaves well under criterion *b*), when considering the first criterion we note that it assumes a constant variance of the errors, so that it is less satisfactory than the GLM2, which allows for a non–constant variance, as outlined in section 3.4.1. In conclusion, graduation method GLM2 performs reasonably well under both criteria in the case of the P model.

As described in section 4, the goodness of fit tests are applied to the Pearson (i.e. standardized) residuals, offering a unified way of comparing various models. As an example, we examine the resulting diagnostic plots for the same groups of observation years as in Figures 3 and 4. In Figures 9 and 10, we consider the diagnostic plots for the P model fitted by the WNLM and GLM2 methods. The plots of residuals against age and the fitted values should show no systematic variation and for a standard normal distribution $N(0, 1)$ the QQ Normal plot should be a straight line. These results are slightly better for the GLM2 models, although there is visual evidence of systematic patterns in the plots against age and fitted values. In the case of the response residuals for the WNLM model, the QQ plots indicate some violations of the model assumptions. Thus, we find similar outcomes when

Comparison of the Parameter Estimates resulted from the main models / Females Duration 1+

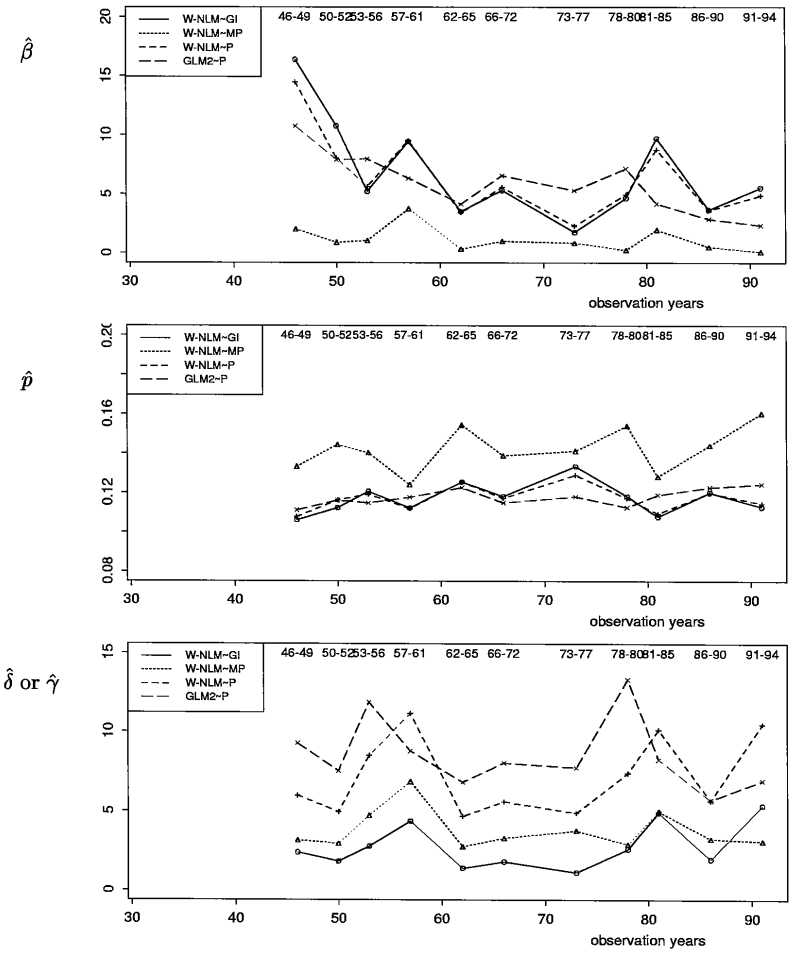

Figure 5: Gompertz and Frailty distribution parameters for Female Annuitants Duration 1+ (age range of 40 – 100+).

Comparison of the Parameter Estimates resulted from the main models / Females Duration 5+.

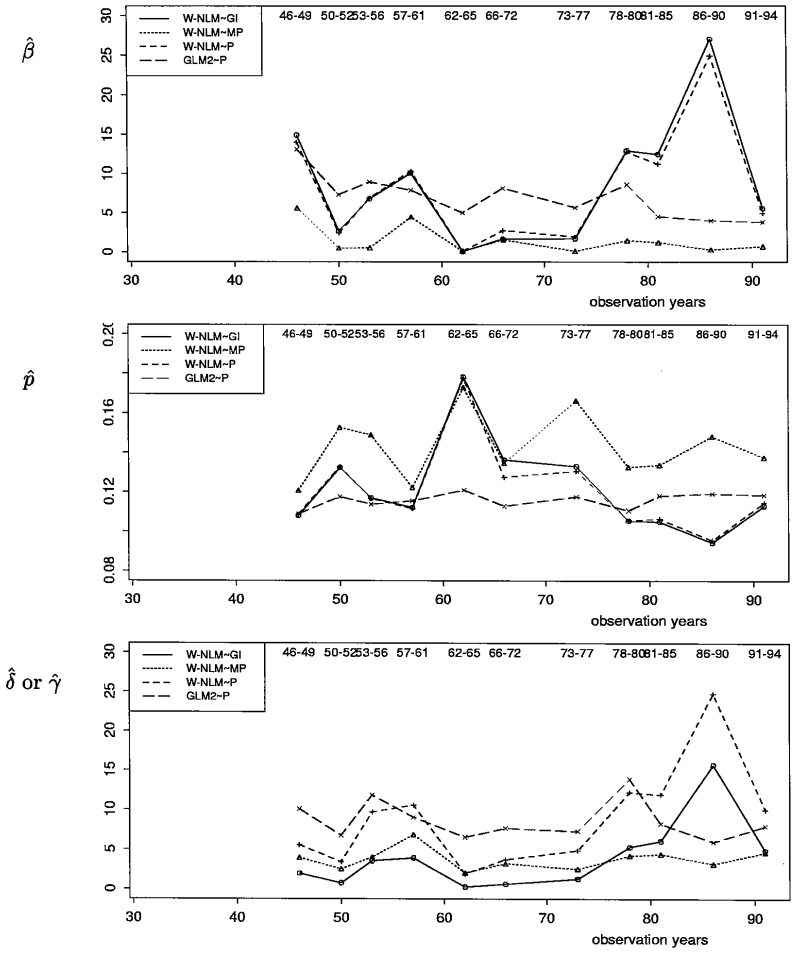

Figure 6: Gompertz and Frailty distribution parameters for Female Annuitants Duration 5+ (age range of 40 – 100+).

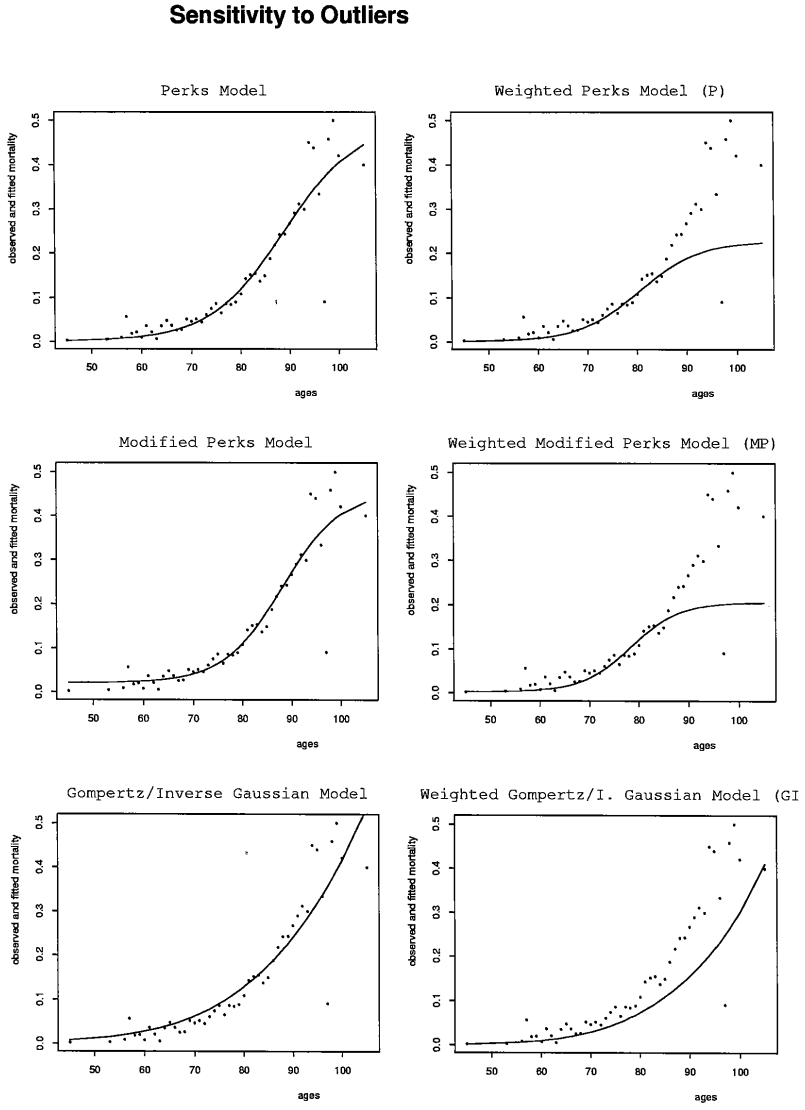

Figure 7: WNLM fitting of Perks, Modified Perks and Gompertz–inverse Gaussian models for Male Annuitants Duration 5+ 1953–56.

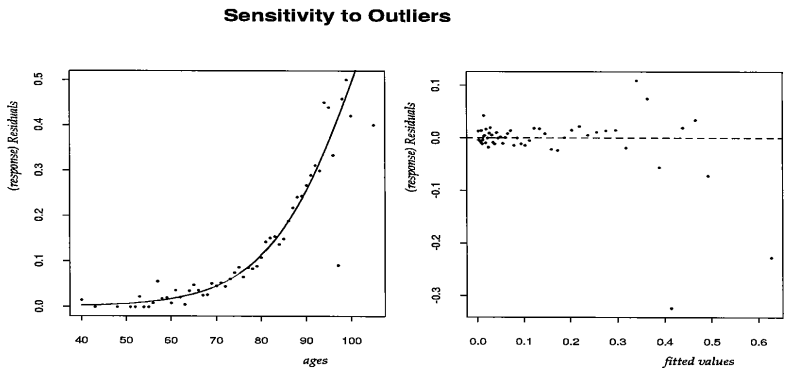

Figure 8: GLM2 fitting of Perks model for Male Annuitants Duration 5+ 1953–56.

comparing the detailed goodness of fit results for the WNLM-P and GLM2-P models, fitted to the different time periods. In the case of the WNLM-P model, the $P(runs)$ and $P(pos)$ statistics take extreme values (i.e. either too large or too small positive numbers of residuals or runs) for most of the graduations whereas in the case of GLM2-P, these statistics take fewer extreme values.

Taking the above considerations into account, we conclude that the GLM method has a stronger theoretical justification and yields models with more favourable properties than the classical non–linear least squares method. The difficulties which have appeared relate to the non–linear parameters in the model (for example, a). Thus, in applications of the GLM2-P model, in some isolated cases, we find that:

a) the minimum value of the deviance curve (see for example, case 1 in Figure 11) lies below the lower bound for parameter a (the bound, which is imposed by the parameterized link function $g(\cdot)$ and the observed forces of mortality — see section 3.4.4); or

b) the deviance curve tends asymptotically to a minimum only when a increases from the lower bound to infinity (see cases 2–4, Figure 11), instead of having a single global minimum for finite values of a (as in all the cases of Figure 1); and we note that

c) there is no straightforward method of estimating the standard error for the parameter a, using the S–PLUS package.

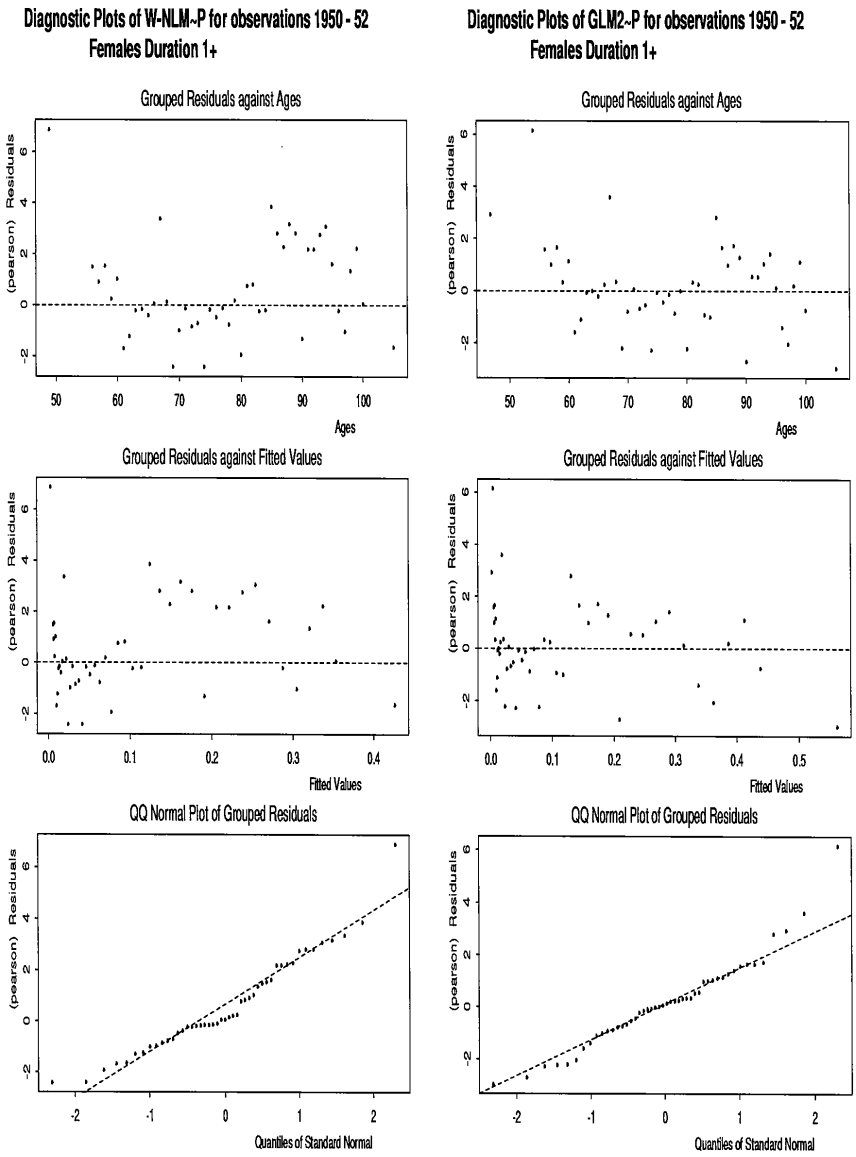

Figure 9: Comparison of WNLM and GLM2 fitting of Perks model for Female Annuitants Duration 1+ 1950 – 52.

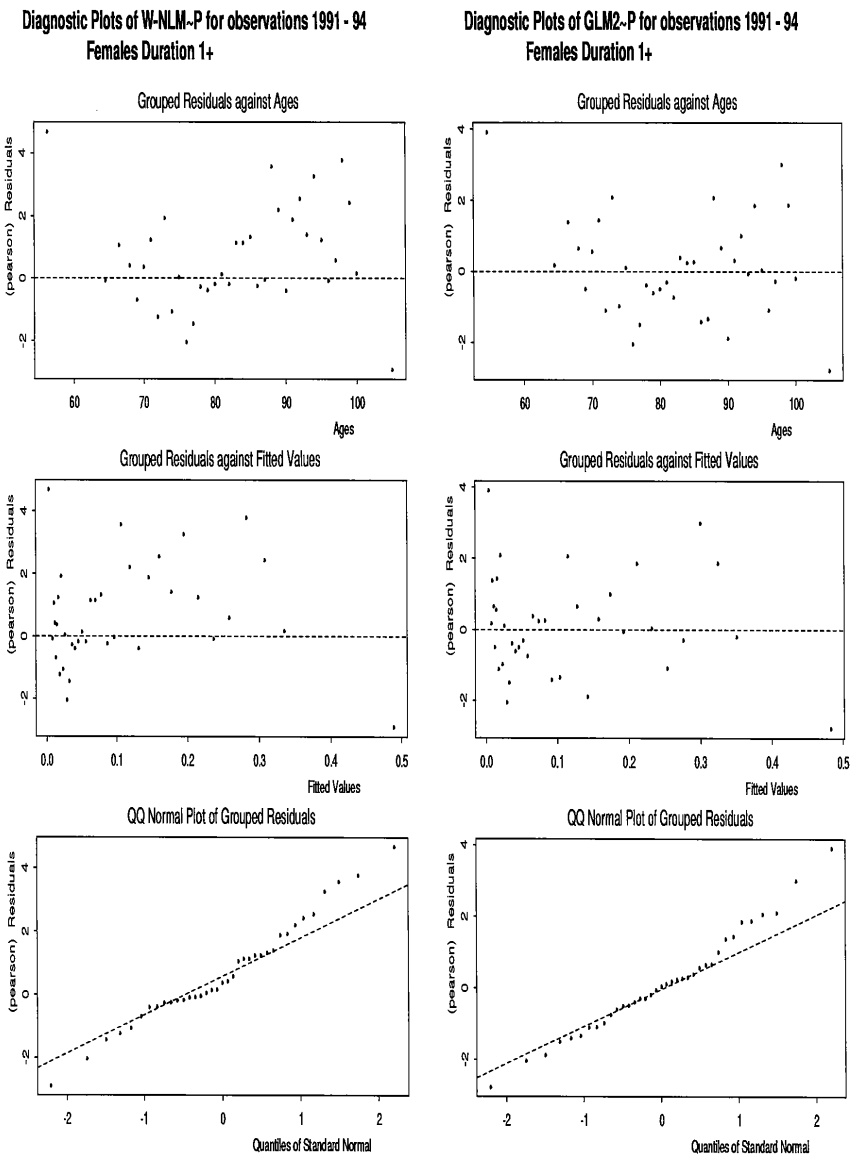

Figure 10: Comparison of WNLM and GLM2 fitting of Perks model for Female Annuitants Duration 1+ 1991 – 94.

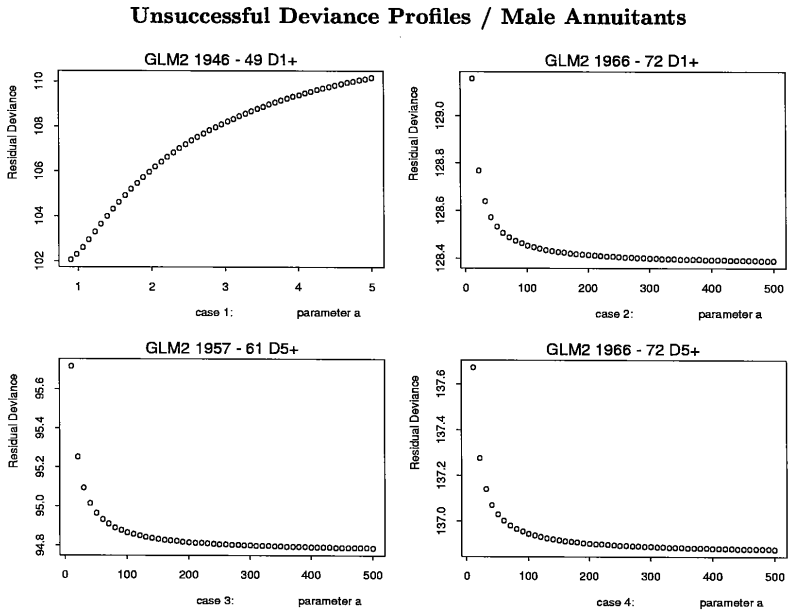

Figure 11: Difficulty with the non–linear parameter in (GLM2) fitting, either $\min(a) < \max(\mu_x)$ or $Dev \to \mathbf{min}$ only if $a \to \infty$.

In cases *a)* and *b)* above, we note that application of the GLM approach with a Taylor Expansion (see section 3.4.3) would provide guidance to the value of $\hat{a}$.

In the last part of the analysis, now that the most suitable model and graduation method have been chosen, in the form of GLM2-P, we proceed to examine further the sensitivity of the model to the choice of age range and presence of outliers. Thus, initially, we restrict the age ranges first to 55 – 100+ and then to 60 – 100+, keeping all the data corresponding to the chosen age range and comparing the graduation results from the intermediate stages. Following that, we investigate the results from fitting the model to the final restricted age range, that is 60 – 100, with any outliers excluded. We note that, deciding whether an observation of a particular data set is an outlier or not depends on the elimination criteria. Thus, when identifying the outliers in a data set, one normally considers the *'distance'* from a given observation to the nearest 'group' of observations relative to the *'mean'* for that particular group. However, in practice, a number of arbitrary factors are involved, making this part

Sensitivity of GLM2-P parameter estimates to outliers
Female Annuitants Duration 1+

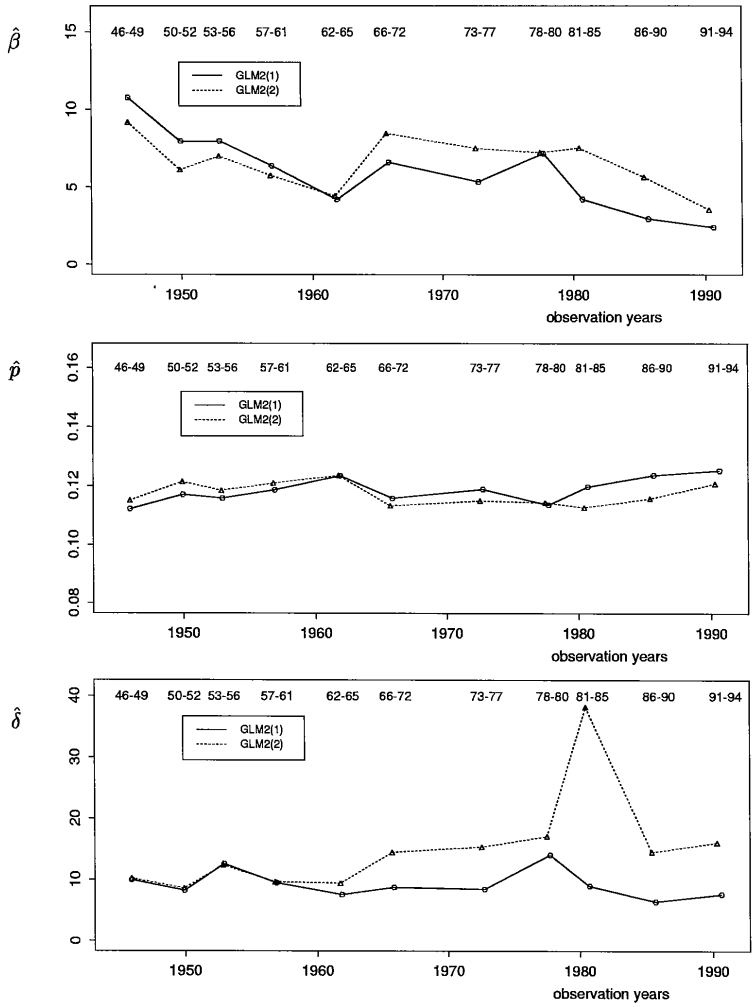

Figure 12: Results of GLM2-P model fitted to age range 60 – 100 applied to data sets with(1) and without(2) outliers.

Sensitivity of GLM2-P parameter estimates to outliers
Female Annuitants Duration 5+

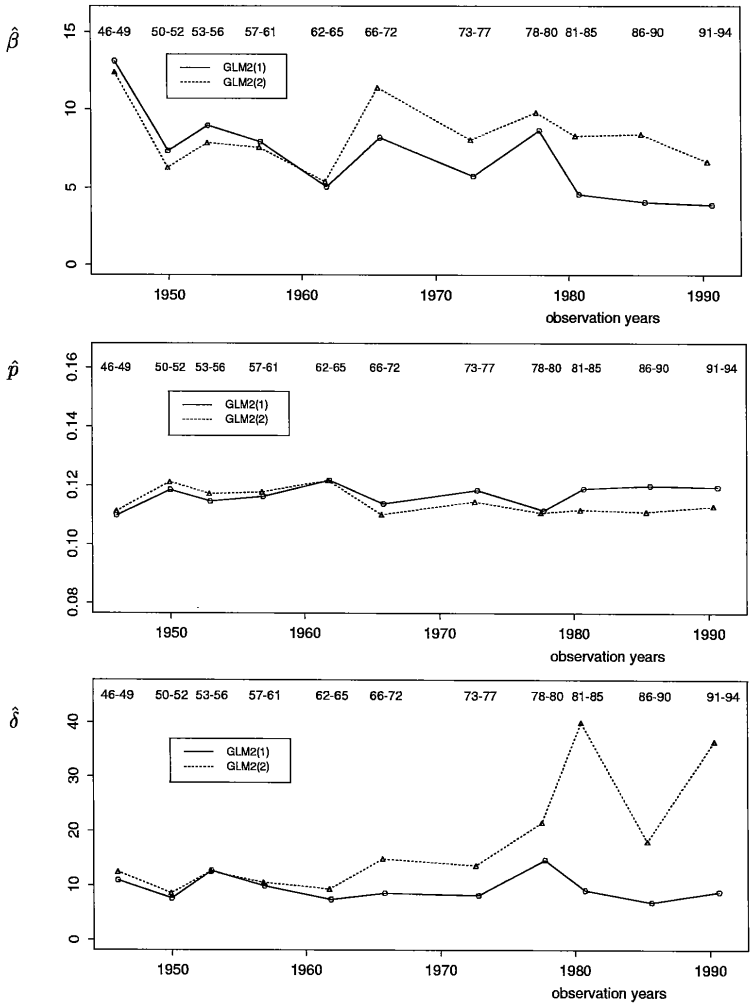

Figure 13: Results of GLM2-P model fitted to age range 60 – 100 applied to data sets with(1) and without(2) outliers.

PERKS MODEL $\left(\frac{a}{1+e^{b-p\cdot x^r}}\right)$ FEMALES DURATION 1+ / GLM2 Fitting Method					
Year	$\hat{a}$	$\hat{b}$	$\hat{p}$	$\hat{\beta}$ $(\times 10^{-6})$	$\hat{\delta} = \hat{\theta}$
1946 – 49	0.99943	7.07184 0.07347	0.11332 0.00182	9.11841	8.81953
1950 – 52	0.85481	7.08636 0.09011	0.11965 0.00221	5.96766	7.14441
1953 – 56	1.28268	7.46787 0.09395	0.11670 0.00223	6.87945	10.99116
1957 – 61	0.97831	7.29746 0.07627	0.11921 0.00179	5.62819	8.20674
1962 – 65	0.97837	7.46792 0.08154	0.12188 0.00190	4.26516	8.02725
1966 – 72	1.45927	7.61223 0.07864	0.11144 0.00180	8.36046	13.09438
1973 – 77	1.58409	7.74673 0.08013	0.11325 0.00180	7.38080	13.98782
1978 – 80	1.76966	7.92022 0.10283	0.11255 0.00227	7.12858	15.72345
1981 – 85	4.12233	8.79027 0.06841	0.11096 0.00149	7.41372	37.15218
1986 – 90	1.49992	7.95078 0.09929	0.11404 0.00212	5.52123	13.15276
1991 – 94	1.74939	8.38516 0.13433	0.11911 0.00278	3.40468	14.68700

Table 1: Regression on Female Annuitants Duration 1+.
Age range: 60 – 100 (outliers removed).

of the analysis more informal in nature. Naturally, there are some significant effects resulting from the elimination of outliers, as illustrated in the following paragraph; however, these effects are less dramatic than those offered by NLM or WNLM models in relation to the same conditions.

In the case of the annuitants' experience, restricting the age range means, in effect, the elimination of most of the 'noise' at the extremes (in the case of the annuitants' experience, this corresponds to observations which are mainly available only for grouped ages, like 40 – 45, 45 – 50 or 100+). This change has little or no effect, for the female annuitants' data in terms of the fitted values, the goodness of fit tests, and the frailty distribution parameter $\hat{\delta}$. This can be observed when comparing the corresponding curves from Figures 5–6 and Figures 12–13. However,

regarding the sensitivity to the outliers, we can note that there are not unexpectedly some changes in the model parameter estimates when graduating with the outliers removed. Although these have very little effect on the predicted values — and we observe improvements in the goodness of fit tests — we note that the effect is concentrated on the $\hat{\beta}$ and $\hat{\delta}$ estimates, as can be seen in Figures 12–13. We also note an upward trend over time for the frailty shape parameter $\hat{\delta}$ for both duration periods 1+ and 5+.

Tables 1 and 2 present the parameter estimates for the female experience for the Perks model fitted by the GLM2 method for the cases of duration 1+ and 5+ respectively (for the age range 60 – 100 with outliers removed).

PERKS MODEL $\left(\frac{a}{1+e^{b-p\cdot x'}}\right)$ FEMALES DURATION 5+ / GLM2 Fitting Method					
Year	$\hat{a}$	$\hat{b}$	$\hat{p}$	$\hat{\beta}$ $(\times 10^{-6})$	$\hat{\delta} = \hat{\theta}$
1946 – 49	1.19455	7.09864 0.07861	0.10952 0.00192	12.35409	10.90756
1950 – 52	0.82905	7.02472 0.10698	0.11961 0.00259	6.16433	6.93107
1953 – 56	1.26636	7.38301 0.09880	0.11551 0.00233	7.75517	10.96351
1957 – 61	1.03058	7.19295 0.07924	0.11607 0.00184	7.46251	8.87899
1962 – 65	0.92482	7.27678 0.10068	0.12010 0.00231	5.24073	7.70026
1966 – 72	1.43285	7.41784 0.08804	0.10825 0.00198	11.32951	13.23674
1973 – 77	1.35996	7.53921 0.09331	0.11278 0.00208	7.94660	12.05903
1978 – 80	2.18555	7.96696 0.10408	0.10893 0.00228	9.71238	20.06465
1981 – 85	4.26086	8.76577 0.07829	0.10989 0.00169	8.19411	38.77211
1986 – 90	1.80723	7.92062 0.11044	0.10924 0.00234	8.30729	16.54385
1991 – 94	3.91063	8.85662 0.15455	0.11109 0.00315	6.54790	35.20359

Table 2: Regression on Female Annuitants Duration 5+.
Age range: 60 – 100 (outliers removed).

6.2 Assured Lives Experience

In the light of the discussion in section 6.1, we focus on the application of the earlier identified model and graduation method of GLM2-P to the assured lives data set. These data are particularly extensive for ultimate durations 5+ and are of good quality. Since the graduation results for the lower durations (1, 2–4) turned out to be inconclusive (for all the methods and models), we present the results here only for the ultimate durations 5+. We also note that we have limited the graduation to age ranges 50 – 100+ in order to allow comparison of results with those obtained in section 6.1 (although, it could also be worthwhile to examine the parameter $\hat{\alpha}$ when the MP model is fitted over the full age range).

When applied to the initial age range of 50 – 100+, the graduations are generally stable and yield parameters which are significantly different from zero throughout the whole data sets. It is important to note, however, that the goodness of fit tests do not support all of the fitted models. Thus the runs and serial correlation tests (which are not independent) show weak graduations for some cases, for example for the periods 1953 – 58, 1987 – 90 and 1991 – 94. The tests indicate strong dependence between individual observations and this is also highlighted by the diagnostic plots in Figure 14, suggesting cyclic patterns in the residuals. Further we note that, for the last two group of observations, the estimates of $\hat{a}$ are very low, 0.253 and 0.237 respectively, suggesting that the models are unsuitable in the light of the interpretation of this parameter as the upper bound for the force of mortality (see section 7.2).

However, when we successively restrict the age ranges (to 55 – 100+ and 60 – 100), automatically removing some of the 'noise' from the data at the older ages, the $\hat{a}$ estimates generally improve together with the goodness of fit tests, except for the 1987 – 90 data which yield a value of $P(runs)$ which is close to zero. Finally, after we remove all possible outliers and we graduate the age range 60 – 100, we note an overall improvement in the graduations for most calendar year groups. The exceptions are again the last two periods mentioned above, where although the estimates of $\hat{a}$ increase to 0.424 and 0.391 respectively, the graduations still produce poor fits, judged by the goodness of fit tests. Figure 15 presents the trends in the parameter estimates arising from the above sensitivity tests for the male assured lives. It is interesting to note that the frailty parameter $\hat{\delta}$ on average is at a lower level compared to the results from the female annuitants data sets (see Figures 5 and 6), and its profile as a function of time is peaked with a maximum in the mid 1950's and a subsequent decline.

Tables 3 and 4 present the parameter estimates for the Perks model fitted by the GLM2 method for the case of ages 50 – 100+ and 60 – 100 (with outliers removed) respectively. In each case the ultimate experience (duration 5+) is considered.

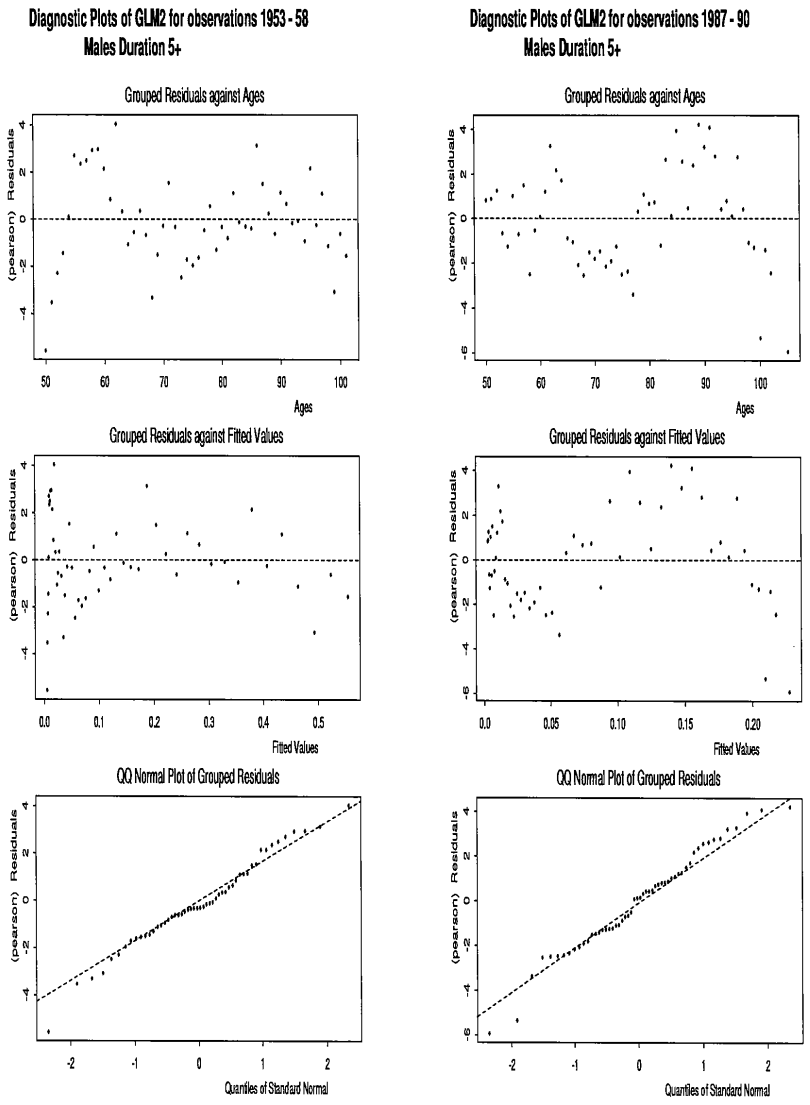

Figure 14: Graduation results of GLM2-P model for Male Assured Lives Duration 5+, age range 50 – 100+

Sensitivity of GLM2-P parameter estimates to outliers
Male Assured Lives Duration 5+

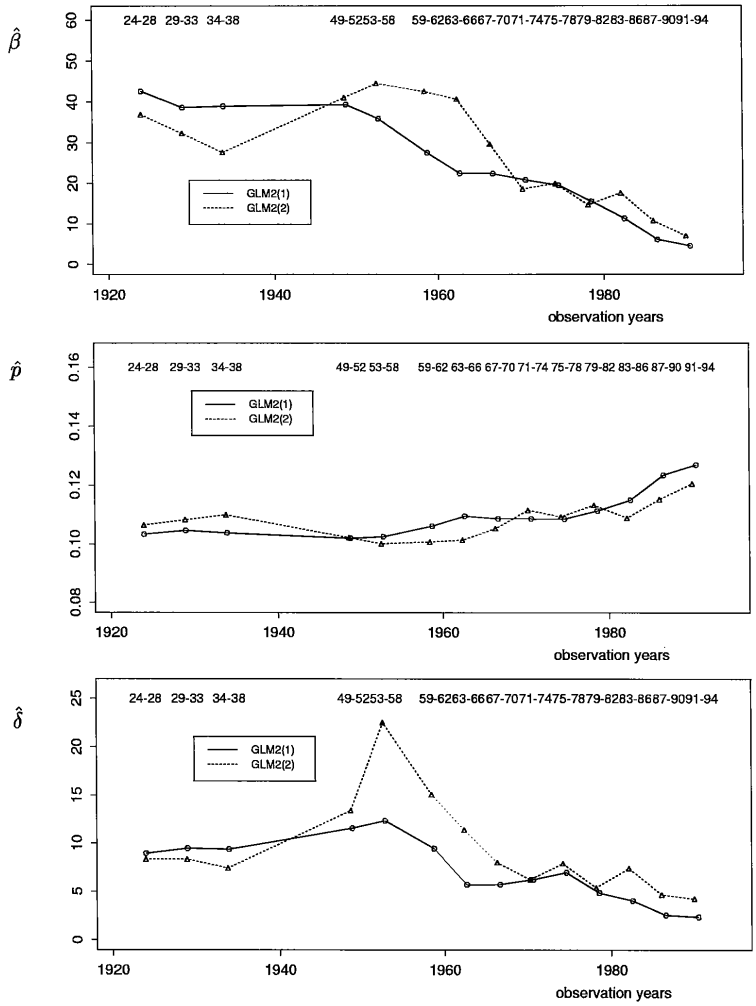

Figure 15: Results of GLM2-P model fitted to age range 60 – 100 applied to data sets with(1) and without(2) outliers.

PERKS MODEL $\left(\frac{a}{1+e^{b-p\cdot x^r}}\right)$ MALES DURATION 5+ / GLM2 Fitting Method					
Year	$\hat{a}$	$\hat{b}$	$\hat{p}$	$\hat{\beta}$ $(\times 10^{-6})$	$\hat{\delta} = \hat{\theta}$
1924 – 28	0.86759	5.83007 0.01879	0.10239 0.00064	42.42544	8.47341
1929 – 33	0.93248	5.94862 0.01850	0.10369 0.00062	38.44513	8.99271
1934 – 38	0.91670	5.95725 0.01781	0.10283 0.00061	38.77908	8.91439
1949 – 52	1.12394	6.22537 0.01454	0.10096 0.00051	39.19496	11.13251
1953 – 58	1.21226	6.36991 0.01549	0.10159 0.00056	35.67910	11.93314
1959 – 62	0.94837	6.24712 0.02059	0.10525 0.00077	27.26334	9.01105
1963 – 66	0.56760	5.80394 0.01778	0.10868 0.00071	22.15022	5.22249
1967 – 70	0.56721	5.84387 0.01588	0.10777 0.00064	22.06230	5.26325
1971 – 74	0.62015	6.00645 0.01421	0.10779 0.00057	20.48702	5.75351
1975 – 78	0.69907	6.19370 0.01482	0.10767 0.00060	19.24066	6.49277
1979 – 82	0.48202	5.94364 0.01798	0.11042 0.00074	15.25889	4.36521
1983 – 86	0.40834	5.96262 0.02300	0.11405 0.00094	10.96962	3.58033
1987 – 90	0.25285	5.78147 0.02913	0.12255 0.00120	5.79592	2.06326
1991 – 94	0.23686	5.91803 0.03597	0.12604 0.00143	4.11950	1.87932

Table 3: Regression on Assured Lives Males Duration 5+.
Age range 50 – 100+.

7 Interpretation of Results

7.1 Identifiability

There are difficulties with interpreting the results of the model fitting exercise described in previous sections and interpreting the suitability of, say, the Perks model

for describing mortality variation with age.

There is an identifiability problem. As noted by Hougaard [1984],

> "The frailty distribution is not identifiable, if the frailty is an individual quantity. This is similar to the normal variance components model. It

PERKS MODEL $\left(\frac{a}{1+e^{b-p\cdot x^{\gamma}}}\right)$ MALES DURATION 5+ / GLM2 Fitting Method					
Year	$\hat{a}$	$\hat{b}$	$\hat{p}$	$\hat{\beta}$ $(\times 10^{-6})$	$\hat{\delta} = \hat{\theta}$
1924 – 28	0.77550	5.77756 0.03042	0.10462 0.00093	36.55031	7.41225
1929 – 33	0.79153	5.85832 0.02811	0.10654 0.00085	31.87819	7.42969
1934 – 38	0.70192	5.83484 0.02980	0.10813 0.00091	27.15500	6.49156
1949 – 52	1.26340	6.32905 0.02567	0.10032 0.00078	40.75522	12.59415
1953 – 58	2.14311	6.86109 0.02006	0.09820 0.00060	44.20519	21.82492
1959 – 62	1.41532	6.46350 0.02837	0.09890 0.00086	42.24255	14.31092
1963 – 66	1.04986	6.18809 0.01929	0.09945 0.00061	40.36880	10.55669
1967 – 70	0.73453	5.99501 0.02445	0.10343 0.00081	29.21118	7.10143
1971 – 74	0.58310	5.99825 0.02441	0.10968 0.00082	18.00226	5.31618
1975 – 78	0.75112	6.26458 0.02422	0.10747 0.00080	19.41134	6.98900
1979 – 82	0.49955	6.01836 0.03154	0.11142 0.00107	14.10104	4.48346
1983 – 86	0.69463	6.33185 0.03150	0.10706 0.00104	17.06057	6.48810
1987 – 90	0.42411	6.10275 0.03252	0.11346 0.00108	10.13939	3.73787
1991 – 94	0.39061	6.27718 0.03380	0.11877 0.00110	6.34391	3.28878

Table 4: Regression on Assured Lives Males Duration 5+.
Age range 60 – 100 (outliers removed).

is not possible to divide the variation into that within and that between individuals, if there is only one observation per individual. It is possible if the frailty is common to several individuals, for example in a family."

Yashin et al [1994] demonstrate that mortality data on their own cannot definitively be used to understand the underlying mechanism of survival. They show that, without additional covariates or assumptions, the fixed frailty Gompertz/ Makeham – gamma model cannot be distinguished from a range of models incorporating changing frailty at the individual level (for example, Le Bras [1976]). Thus, "even when a model fits the data, the concepts used to construct the model may not be correct".

This conclusion follows on from Hoem [1990] who proved that many underlying mortality rates and many frailty distributions produce the same mortality pattern. As a specific example, we consider the Gompertz – gamma model of (A.8)

$$\overline{\mu}(x) = \frac{\delta}{\theta + H_x}\,\mu_x \tag{A.8}$$

where $\mu_x = \beta \cdot e^{p \cdot x}$ and $H_x = \int_0^x \mu_t \,\mathrm{d}t$.

Then, given a non–negative random variable with distribution function $G(t)$ and finite mean, it is possible to construct a new frailty model with individual force of mortality (corresponding to $Z = 1$) ν_x and frailty distributed at birth with $f_0(z) = G'(z)$ such that the population hazard rate matches (A.8) exactly. The form of ν_x is:

$$\nu_x = \frac{\delta \cdot \mu_x}{\theta + H_x}\, b\left(\delta \cdot \ln\left(1 + \frac{H_x}{\theta}\right)\right) \tag{7.1}$$

where $b(\cdot)$ is a function defined by the following steps:

$$\begin{aligned} r(x) &= \int_0^\infty u \cdot e^{-u \cdot x}\,\mathrm{d}G(u) \;\Big/\; \int_0^\infty e^{-u \cdot x}\,\mathrm{d}G(u) \\ R(x) &= \int_0^x r(y)\,\mathrm{d}y \\ B(x) &= R^{-1}(x) \\ b(x) &= \frac{\mathrm{d}}{\mathrm{d}x} B(x)\,. \end{aligned}$$

7.2 Interpretation of Parameters

For the Perks model, we can write

$$\overline{\mu}(x) = \frac{a}{1 + e^{b - p \cdot x}}$$

and we can interpret the parameters as follow:

1. a is clearly the limiting asymptotic value for the population force of mortality:

$$a = \lim_{x \to \infty} \overline{\mu}(x) .$$

2. In the Gompertz model

$$\mu_x = \beta \cdot e^{p \cdot x}$$

and it is clear that p measures the relative rate of increase in the force of mortality with age. For the Perks model, we note that, following Thatcher [1999]:

$$\lambda(x) = \frac{1}{\overline{\mu}(x)} \frac{\mathrm{d}}{\mathrm{d}x} \overline{\mu}(x) = p \left(1 - \frac{\overline{\mu}(x)}{a} \right) . \tag{7.2}$$

Thus, the relative rate of increase is close to p at young ages, where $\overline{\mu}(x)$ is small, but then it reduces as age increases and for example at ages high enough for $\overline{\mu}(x)$ to be approximately $\frac{1}{2}a$, the relative rate of increase would have fallen to $\frac{1}{2}p$.

3. b can be interpreted via the age x_0 where

$$x_0 = 40 + \frac{b}{p}$$

which is the age at which $\overline{\mu}(x) = \frac{1}{2}a$, i.e. half its asymptotic value, and the age at which the Perks curve has its point of inflection.

7.3 Comparison of Results

It will be useful to compare the result with those obtained from other similar studies.

Estimates of p are available from a number of recent studies. Thus, Forfar and Smith [1988] fit the full Heligman – Pollard model to successive English Life Tables. At the oldest ages, this model reduces to

$$q_x = \frac{G \cdot H^x}{1 + G \cdot H^x} \tag{7.3}$$

and so $\ln H$ is directly comparable to p in our notation, setting aside the difference between q_x and μ_x. For ELTs No 11 (1950 – 52), 12 (1970 – 72) the estimates for $\hat{p}$ would be 0.10003, 0.10346 and 0.10373 respectively. However, we note that this exercise involved fitting to the already graduated q values rather than to the crude data.

Congdon [1993] uses adult mortality data (ages 25 – 90) for Greater London for 1980 – 82 and fits a range of models, including

a) one based on the ELT 12 graduation formula and

b) the modified Heligman – Pollard model

$$q_x = \frac{G \cdot H^x}{1 + k \cdot G \cdot H^x} \tag{7.4}$$

For the parameters corresponding to p, the estimates are

a) 0.0716 and 0.1086 for males and females respectively

b) 0.10400 and 0.09685 for males and females respectively.

(We note that in the case of *b)* the estimates of $\hat{k}$ are negative for females: as discussed also in the **Appendix**).

Wang and Brown [1998] fit the Gompertz – gamma model to the 1994 Group Annuity Reserving Table of the Society of Actuaries, obtaining the following estimates, in our notation:

$$\hat{a} = 0.15359 \quad \hat{b} = 5.20153 \quad \hat{p} = 0.11761 \quad \hat{\beta} = 7.762 \times 10^{-6} \quad \hat{\theta} = 1.306$$

Given the interpretation of a as $\lim_{x \to \infty} \overline{\mu}(x)$, this particular estimate $\hat{a}$ looks very low!

Manton et al [1986] fit the Gompertz – gamma model to US Medicare data for both sexes by cohort during the period 1968 – 78. Their $\hat{p}$ estimates for cohorts born between 1884 and 1902 lie in the range 0.0734 – 0.1029. The θ estimates are published for the 1892 birth cohort (ages 75 – 85):

$$\begin{cases} \hat{\theta} = 4.7393 & \text{for males} \\ \hat{\theta} = 3.4722 & \text{for females} \end{cases}$$

Congdon [1994] presents a theoretical argument to demonstrate that, all other parameters remaining unchanged, an increase in the gamma variance (i.e. a decrease in θ) is associated with an increase in life expectancy. The detailed results of Tables 1–4 (and unpublished results) and Figures 5, 6, 12, 13 and 15 are not conclusive in this regard because the other parameter estimates (p and β representing level and slope of mortality at the individual level) are different as we move from one period to the next. However, Congdon [1994] further reports a tendency for the gamma variance to reduce over time (i.e. corresponding to an increase in θ) in a series of fits of frailty models (with covariates) to male mortality (at ages 25 – 90) for Inner and Greater London over the period 1971 – 90. The ranges for the parameters are as follows:

$$\begin{cases} \hat{a} : & 0.4300 & — & 9.7299 \\ \hat{p} : & 0.0968 & — & 0.1075 \\ \hat{\beta} : & 26.9912 \times 10^{-6} & — & 40.7930 \times 10^{-6} \\ \hat{\theta} : & 3.9988 & — & 100.4842 \end{cases}$$

Congdon [1994] also uses the Gompertz – inverse Gaussian model and finds a slight improvement in fit in relation to the Gompertz – gamma model: this can be compared with the opposite finding of the study of Manton et al [1986] and this report. Further, Congdon experiments with Aalen's more general model (A.25) and finds estimates of $\hat{\psi}$ that are closer to 2 than 1, indicating the slight superiority of the inverse Gaussian frailty distribution in this case.

Damaskos [1998] uses the Gompertz – gamma model for recent English Life Tables fitting the models by the weighted non–linear squares method to the underlying crude data (unlike Forfar and Smith [1988]) The parameter estimates are presented in Table 5, using a range of adult ages. The ELT 15 results for females are not included because of the poor fit provided by the model.

Parameter Estimates					
Period	Age range	$\hat{a}$	$\hat{p}$	$\hat{\beta}$ ($\times 10^{-6}$)	$\hat{\theta}$
MALES					
ELT 12	30 – 90	1.1384	0.0964	66.603	11.8170
	40 – 90	1.1506	0.0962	67.473	11.9640
	50 – 90	1.2111	0.0953	71.801	12.7065
ELT 13	30 – 90	0.0852	0.0959	67.448	8.8790
	40 – 90	0.0863	0.0957	68.753	9.0148
	50 – 90	0.0900	0.0948	73.400	9.5008
ELT 14	30 – 90	0.0669	0.1042	33.393	6.4219
	40 – 90	0.0671	0.1042	33.583	6.4430
	50 – 90	0.0679	0.1038	34.490	6.5406
ELT 15	30 – 90	0.0748	0.1035	28.318	7.2258
	40 – 90	0.0747	0.1035	28.260	7.2147
	50 – 90	0.0759	0.1031	29.042	7.3619
FEMALES					
ELT 12	30 – 90	0.0889	0.1172	8.576	7.5839
	40 – 90	0.0882	0.1174	8.470	7.5169
	50 – 90	0.0855	0.1181	8.008	7.2386
ELT 13	30 – 90	0.0989	0.1122	10.979	8.8140
	40 – 90	0.0980	0.1123	10.870	8.7310
	50 – 90	0.0933	0.1132	10.197	8.2460
ELT 14	30 – 90	3.017	0.1089	11.814	27.7010
	40 – 90	2.952	0.1089	11.734	27.1100
	50 – 90	2.644	0.1095	11.273	24.1500

Table 5: Parameter Estimates from Damaskos [1998]

7.4 Forecasting Parameter Values

Keyfitz [1982] argues persuasively that the effective forecasting of mortality rates (and other demographic variables) depends on finding a parameter space of small dimension in which the mortality pattern for the population under consideration can be located. For forecasting, we would like the successive sets of mortality rates over time to be represented by point moving in a simple way through the parameter space — ideally in a straight line (see Sithole et al [2000] for an example). This is a two stage process. Finding an appropriate parameter space representing a good fit to past set of data will not be sufficient — the second requirement is a simple and predictable progression through time. The analysis so far has only addressed the first stage and our favoured Gompertz - gamma model provides an adequate representation through three parameters (a, b and p or β, p and θ).

Given the trends depicted in section 6 for the principal parameters in the Gompertz - gamma model, the next step is to forecast future values for the parameters. This is the approach first advocated by Cramer and Wold [1935] in their modelling of Swedish mortality over time from 1801 - 1930 via the Makeham formula

$$\overline{\mu}(x, y) = \alpha_y + \beta_y \cdot c_y^x \tag{7.5}$$

where α_y, β_y, c_y were modelled by parametric formulae involving calendar time y. Specifically α_y was represented by a linear function and $\log \beta_y$ and $\log c_y$ each by a logistic function.

An alternative methodology would be to use time series methods of the Box–Jenkins type (see McNown and Rogers [1989]) or the structural time series methods of Harvey and Durbin [1986].

Wang and Brown [1998] take a different approach and propose a model where the overall improvement in mortality rates is proportional to the average frailty in the population i.e. they define the annual improvement factor

$$\overline{E}_{x,y} = 1 - \frac{\overline{\mu}(x, y+1)}{\overline{\mu}(x, y)}$$

and then suggest

$$\overline{E}_{x,y} = \kappa \cdot \overline{z}_x \tag{7.6}$$

where κ is a parameter to be estimated from past data and $\overline{z}_x$ is given by (A.7) for the Gompertz - gamma model. This proposal would imply that as average frailty decreases with age so the annual improvement factors would also decrease. It would enable the forecasting of 1 year ahead values i.e. $\overline{\mu}(x, y+1)$ from knowledge of $\overline{\mu}(x, y)$. But the implications are that

$$\begin{aligned} \overline{\mu}(x, y+1) &= \overline{\mu}(x, y)\left(1 - \overline{E}_{x,y}\right) \\ &= \overline{\mu}(x, y)\left(1 - \kappa \cdot \overline{z}_x\right) \end{aligned}$$

which is *not* of the Gompertz – gamma structure and is unlikely to produce a good fit to the data as they emerge. Indeed, it is straightforward to show that, in this case, we would have

$$\overline{\mu}(x, y+1) = \frac{\sum_{n=0}^{1} \alpha_n \cdot e^{p \cdot x(n+1)}}{\sum_{n=0}^{2} \beta_n \cdot e^{p \cdot x(n+1)}}$$

for suitably chosen α_0, α_1, β_0, β_1, β_2.

An alternative approach which has not been investigated further would be to begin with a model of the form

$$\mu(x, y, z) = f(y, z) \cdot \mu(x, 1, 1) = f(y, z) \cdot \mu_x \tag{7.7}$$

and consider simple choices of $f(\cdot)$, for example

$$f(y, z) = f_0 + f_1^y \cdot z \tag{7.8}$$

where f_1^y is different for each calendar period y and $f_1^y > f_1^{y+1}$, and the main mortality structure remains fixed over time. Then, in an obvious notation,

$$\begin{aligned} E_{x,y,z} &= 1 - \frac{\mu(x, y+1, z)}{\mu(x, y, z)} \\ &= \frac{\left(f_1^y - f_1^{y+1}\right) z}{f_0 + f_1^y \cdot z} \end{aligned} \tag{7.9}$$

which implies that improvement would be higher for the more frail individual.

This extension of our analysis remains for future work.

In attempting to understand past mortality trends and consider forecasts for the future, it is important to begin with an accurate and correct assessment of the past. As many investigators have concluded, including for example Vaupel et al [1979] and Congdon [1994], the presence of heterogeneity casts doubt on conventional methods of measurement. It thus becomes important to allow fully for frailty in models of mortality so that trends particularly at the oldest ages can be identified — failure to do so could understate mortality improvement at these ages, in particular (Vaupel et al [1979]).

8 Conclusions

The theoretical implications of heterogeneity (viz frailty in human populations) when estimating individual mortality rates from available population data have been extensively studied in the demographic and statistical literature. Various models have been proposed for measuring the size of such heterogeneity in human mortality data,

based on discrete and continuous age formulations. In this paper, we have considered the latter in the form of applications of Vaupel et al [1979]'s multiplicative model, because it is more straightforward to implement in practical applications and to interpret the results, than the discrete alternatives. In this context, we have investigated, using two extensive insurance–based data sets, two frailty distributions, namely the gamma and the inverse Gaussian with respect to the Gompertz/Makeham family of individual hazard functions. The effects on the outcomes of using different graduation techniques have also been studied.

Thus, regarding the assumptions for the frailty distribution, the analysis shows that the model incorporating the inverse Gaussian frailty distribution (GI) generally results in significant model parameters from the graduations, for the relevant age ranges. But this model is less responsive to variations in the data and also is less likely to lead to satisfactory predictions for human mortality at the older ages (being an exponentially increasing function with age) than alternatives. In contrast, the parameter estimates offered by the models assuming a gamma frailty distribution (MP or P) are less significant, but the models are more favourable because of a better fit and because the model has a logistic form with respect to age. Therefore, we conclude that, when applied together with the Gompertz/Makeham individual hazard functions, the gamma frailty distribution seems to be more viable compared to the inverse Gaussian. It also predicts a non–homogeneous population for older ages, as discussed by Manton and Stallard [1984]. Further, we confirm the experience of other authors, that, when applied to the older ages only, the MP model is inferior to the P model, as the $\hat{\alpha}$ estimates do not appear to be significant.

The generalized linear modelling method has a stronger theoretical justification compared to the classical non–linear least squares method and can result in more stable parameter estimates. However, there are potential drawbacks when applying the GLM fitting technique to models with non–linear parameters. Firstly, one needs to know in advance the most likely domain for the non–linear parameters and secondly, the fitting process can fail to produce a global minimum of the deviance profile for the respective domain. We note that, in the particular case of the P model, the first disadvantage is usually not serious, as we can often estimate permissible intervals for the a parameter. Further, this disadvantage is less serious, when compared with the traditional method of non–linear regression, where we need 'good' starting values for 'all' of the parameters involved. Although the second drawback seems to reduce the chances of success for fitting with the GLM2 method for some of the data sets investigated, it should be noted that, for all of these cases, there were also difficulties with the NLM method. Also, we note that the P models fitted by W-NLM technique proved to be strongly influenced by outliers in the data sets, often yielding, for example, estimates of $\hat{a}$ that are too low compared to the observed forces of mortality (where we recall that $a = \lim_{x\to\infty} \mu_x$).

Regarding the size of the frailty effect for the investigated data sets, the results are less conclusive with a coefficient of variation of the gamma frailty distribution

$(1/\sqrt{\delta})$ for the final models being on average less than 0.45 for the different data sets, implying 'fairly' homogeneous populations (corresponding to a $\hat{\delta}$ value of approximately 5). As has been shown earlier, parameter values corresponding to $\hat{a}$ have been found in the literature in the range 0.15 to 9.5. Considering the interpretation of a, we probably can restrict this range to the interval from 0.35 to 1.3. For the relative rate of increase of the force of mortality $\hat{p}$, the estimates are fairly stable, lying between 0.09 to 0.12. Thus, we could draw up the potential range of the gamma shape parameter, δ, that could normally result from the P model as being:

$$\hat{\delta} = \frac{\hat{a}}{\hat{p}} = \frac{0.35}{0.12} \rightarrow \frac{1.3}{0.09} \simeq 2.916 \rightarrow 14.444 \,.$$

It is possible that the moderately high values of $\hat{\delta}$ are attributable to the investigated populations being shaped by selection (temporary initial selection for assured lives and adverse or self–selection by annuitants: see Benjamin and Pollard [1993]). For such populations the basic frailty models, with constant frailty at the individual level during a life span, might not be sufficiently sensitive to pick up differences between the individual members. Then, a natural extension would be to study more advanced models where frailty varies with other factors, including age and physiological and socio–economic variables.

Overall, we conclude that the fitting of the Gompertz – gamma model (P) has been broadly successful for the two data sets discussed i.e. female annuitants and male assured lives (despite some discrepancies — for example Figure 14). For these two data sets, the detailed results indicate that it would be worth investigating modelling the time trends in these parameter values with a view to developing methods for forecasting using time series methods (as in McNown and Rogers [1989]) or incorporating the time parameter directly in the frailty model (as advocated by Manton et al [1981] and Vaupel [1999]: see equations (2.16) and (2.17)). This will be the subject of future research.

Finally, we return to the identifiability discussion of section 7.1 and the need for caution in interpreting the meaning of the parameters in any frailty–based model.

Acknowledgments

A special thanks should go to our colleague Arthur Renshaw for the invaluable comments and insights into the GLM technique during the preparations of this work. Also the authors are grateful to the CMI Bureau for providing the data.

This research work was funded by a research grant from the Institute and Faculty of Actuaries, and by a research grant from the CMI Bureau.

Appendix:

Mathematical Background: Frailty Models

As noted in section 2.1, two important components of the frailty model are the forms for μ_x and $f_0(z)$. As mentioned in the introduction, we concentrate on the Gompertz – Makeham form for the underlying individual hazard function as it has been found by many studies that this form fits most mortality data well in the adult age range. Thus, we consider the force of mortality for the 'standard' individual:

$$\mu_x = \mu(x|1) = \alpha + \beta \cdot e^{p \cdot x}, \tag{A.1}$$

where α is the Makeham correction term for young age mortality and its value is usually close to or equal to zero.

It is convenient to treat the two possible cases separately, that is:

$\alpha = 0$ **Gompertz Model**

According to our definition in (2.2) the 'standard' cumulative hazard is given by:

$$\begin{aligned} H_x &= \int_0^x \beta \cdot e^{p \cdot t} \, \mathrm{d}t = \frac{\beta}{p} e^{p \cdot t} \Big|_0^x \\ &= \frac{\beta}{p} (e^{p \cdot x} - 1) \, . \end{aligned} \tag{A.2}$$

$\alpha \neq 0$ **Makeham Model**

The definition (2.1) of the individual forces of mortality is modified to $\mu_\alpha(x|z) = \alpha + z \cdot \mu_x$ (2.1α) making the cumulative hazard conditional on z become:

$$\begin{aligned} H_\alpha(x|z) &= \int_0^x \mu(t|z) \, \mathrm{d}t = \int_0^x \alpha \, \mathrm{d}t + z \int_0^x \mu_t \, \mathrm{d}t \\ &= \alpha \cdot x + z \cdot H_x \, , \end{aligned} \tag{2.2α}$$

where H_x is given by (A.2).

Similarly, the expression for the cohort force of mortality will change slightly from (2.9):

$$\begin{aligned} \overline{\mu}_\alpha(x) &= \int_0^\infty \mu(x|z) \cdot f_x(z) \, \mathrm{d}z \\ &= \int_0^\infty (\alpha + z \cdot \mu_x) \, f_x(z) \, \mathrm{d}z \\ &= \alpha + \overline{z}_x \cdot \mu_x \, . \end{aligned} \tag{2.9α}$$

It can be shown that the above changes will have no influence on $f_x(z)$, the *p.d.f.* of frailty for those alive at a current age x. Using the definition (2.8) for this, we obtain:

$$\begin{aligned} f_{x,\alpha}(z) &= \frac{f_0(z)\cdot S_\alpha(x\,|\,z)}{\int_0^\infty S_\alpha(x\,|\,z)\cdot f_0(z)\,\mathrm{d}z} = \frac{f_0(z)\cdot e^{-H_\alpha(x|z)}}{\int_0^\infty f_0(z)\cdot e^{-H_\alpha(x|z)}\,\mathrm{d}z} \\ &= \frac{f_0(z)\cdot e^{-(\alpha\cdot x+z\cdot H_x)}}{\int_0^\infty f_0(z)\,e^{-(\alpha\cdot x+z\cdot H_x)}\,\mathrm{d}z} = \frac{e^{-\alpha\cdot x}\cdot f_0(z)\cdot e^{-z\cdot H_x}}{e^{-\alpha\cdot x}\int_0^\infty f_0(z)\cdot e^{-z\cdot H_x}\,\mathrm{d}z} \\ &= f_x(z)\,. \end{aligned}$$

Hence, the expected value $\mathrm{E}[Z_x] = \overline{z}_x$ is the same as for the Gompertz model. The implication is that, in the Makeham model, the scale parameter α acts additively as an independent term in the case of the population force of mortality function (see (2.9α)), in the same way as the force of mortality for the 'standard' individual (A.1).

However, we note that, since in all of the applications we have focused on the adult and older ages, the inclusion of this correction factor α (initially introduced by Makeham in order to improve the representation of mortality rates at the younger ages), does not improve the goodness of fit so that effectively $\alpha = 0$ throughout.

Regarding the choice of $f_0(z)$, Vaupel et al [1979] have proposed the gamma distribution. Their results were generalized by Hougaard [1984] who showed that if the frailty *p.d.f.* $f_0(z)$ came from the non–negative exponential family with z as a canonical statistic viz

$$P(\delta,\,\theta) = \frac{z^\delta\cdot\exp(-\theta\cdot z)\cdot m(z)}{\varphi(\delta,\,\theta)} \tag{A.3}$$

then the resulting *p.d.f.* $f_x(z)$ among surviving members of the cohort will belong to the same frailty with modified parameters i.e. $P(\delta,\,\theta + H_x)$.

This family includes the degenerate distribution (i.e. homogeneous case), gamma, inverse Gaussian, Poisson, two–point, truncated normal and non–central χ^2 distributions. Hougaard [1986] extended these results to include stable distributions.

We report the empirical results of fitting models based on two types of frailty *p.d.f.* at birth, namely: gamma and inverse Gaussian.

a) **Gamma distribution:**

$$P(\delta,\,\theta) = f_0(z) = \frac{\theta^\delta\cdot z^{\delta-1}}{\Gamma(\delta)}\,e^{-\theta z}\,, \tag{A.4}$$

then the distribution of frailty among survivors at any age x in the cohort is given by:

$$P(\delta,\,\gamma+H_x) = f_x(z) = \frac{(\gamma+H_x)^\delta\cdot z^{\delta-1}}{\Gamma(\delta)}\,e^{-z(\gamma+H_x)}\,. \tag{A.5}$$

Hence, the average frailty at birth and among survivors at any age x respectively will be:

$$\overline{z}_0 = \frac{\delta}{\theta}, \tag{A.6}$$

$$\overline{z}_x = \frac{\delta}{\gamma + H_x}. \tag{A.7}$$

We observe that the shape parameter δ from the above gamma distribution function is an important indicator of the extent of heterogeneity in the population: $f_0(z)$ has coefficient of variation $= 1/\sqrt{\delta}$. Thus, the smaller the value of δ the more significant is the heterogeneity in the population and when $\delta \to \infty$ the coefficient of variation in frailty becomes *zero* resulting in the individual and population hazard rates becoming identical.

Note that we could obtain the same result from the *p.d.f.* of $\boldsymbol{Z}_x$ (2.8) by substituting the expressions for the conditional survivor function $S(x \mid z)$, given by (2.3), and the cumulative hazard (2.2) together with the marginal *p.d.f.* (A.4).

Then substituting the above expression for the expected frailty (A.7) into (2.10) yields the intrinsic relationship between the population hazard and the individual hazard function:

$$\overline{\mu}(x) = \frac{\delta}{\gamma + H_x}\mu_x. \tag{A.8}$$

We can develop the last relation further making use of the cumulative hazard (A.2) and also allowing for the two types of individual hazard functions:

1. **Gompertz Hazard:** $\mu_x = \beta \cdot e^{p \cdot x}$
   Then,

$$\begin{aligned}\overline{\mu}(x) &= \frac{\delta \cdot \beta \cdot e^{p \cdot x}}{\theta + \frac{\beta}{p}(e^{p \cdot x} - 1)} \\ &= \frac{\delta \cdot p \cdot \beta \cdot e^{p \cdot x}}{(\theta \cdot p - \beta) + \beta \cdot e^{p \cdot x}}.\end{aligned} \tag{A.9}$$

   This is one of the principal models investigated in this paper. However, presentation of (A.9) has been transformed in order to make the non–linear numerical fitting to the data more stable. According to Ratkowsky [1983, 1990], these transformations not only make the function more suitable for parameter predictions (better statistical properties of the estimators), but also lead to parameters that have direct meanings. Thus we write:

$$\begin{aligned}\overline{\mu}(x) &= \frac{\delta \cdot p}{1 + \frac{\theta \cdot p - \beta}{\beta} e^{-p \cdot x}} \\ &= \frac{a}{1 + e^{b - p \cdot x^r}},\end{aligned} \tag{A.10}$$

where in the above $x' = x - 40$ is the age variable translated to the minimum observation age in the data, which improves the computations significantly. We assume throughout that the mean frailty $\overline{z}_0 = 1$ (i.e. $\delta = \theta$) in the initial population, so that the mean coincides with the standard value at age zero. Then, we can calculate the estimated parameters of the individual hazard rate function and those of the frailty distribution:

$$\hat{\beta} = \frac{\hat{a}}{1 + e^{\hat{b}+40\,\hat{p}}}, \tag{A.11}$$

$$\hat{\theta} = \hat{\delta} = \frac{\hat{a}}{\hat{p}}, \tag{A.12}$$

where we observe that $\beta = \mu_0 = \overline{\mu}(0)$ as expected when $\overline{z}_0 = 1$.

2. **Makeham Hazard:** $\mu_x = \alpha + \beta \cdot e^{p \cdot x}$

The cohort force of mortality resulting from (2.9α) and (A.8) in this case is

$$\overline{\mu}_\alpha(x) = \alpha + \frac{\delta \cdot p \cdot \beta \cdot e^{p \cdot x}}{(\theta \cdot p - \beta) + \beta \cdot e^{p \cdot x}} \tag{A.9α}$$

which can be transformed into the form:

$$\overline{\mu}_\alpha(x) = \alpha + \frac{a}{1 + e^{b - p \cdot x'}}. \tag{A.13}$$

Both (A.9) and (A.9α) are members of the family of mortality curves first proposed by Perks [1932] for graduating the force of mortality. Beard [1959, 1971] undertook a series of investigations, based on different theoretical models of the mortality process, which could lead to the Perks family representation of the force of mortality. One of these includes the proposal of a heterogeneous population and the derivation of (A.9α) for the Makeham – gamma case. This study therefore pre–dates Vaupel et al [1979] and we note that Manton et al [1986] includes a reference to Beard's work.

In passing, we note that the formula (A.13) has appeared in the actuarial literature, as part of the graduation formula for adult mortality, for E.L.T. Nos. 11 and 12 (Benjamin and Pollard [1993]).

b) **Inverse Gaussian distribution:**

$$f_0(z) = \left[\frac{\psi}{\pi z^3}\right]^{\frac{1}{2}} \exp\left(4\psi\gamma\right)^{\frac{1}{2}} \exp\left(-\gamma z - \psi z^{-1}\right), \quad z > 0 \tag{A.14}$$

which can be re-parameterized, on completing the square, as

$$f_0(z) = \left[\frac{\lambda}{2\pi\gamma^2 z^3}\right]^{\frac{1}{2}} \exp\left[-\frac{(z-\lambda)^2}{2\lambda\gamma^2 z}\right] \tag{A.14$'$}$$

Hougaard [1984] shows that with respect to the standard form for the exponential family, $P(\delta, \gamma)$, $\delta = -\frac{1}{2} = \textbf{const}$ and that the distribution of frailty among survivors at age x will be:

$$f_x(z) = \left[\frac{\psi}{\pi z^3}\right]^{\frac{1}{2}} \exp\left[4\varphi\ (\gamma + H_x)\right]^{\frac{1}{2}} \exp\left[-(\gamma + H_x)\, z - \psi\, z^{-1}\right] . \quad \text{(A.15)}$$

It is straightforward to show that the average frailty at birth and among survivors at age x will be:

$$\overline{z}_0 = \left(\frac{\psi}{\gamma}\right)^{\frac{1}{2}} \quad \text{(A.16)}$$

$$\overline{z}_x = \left(\frac{\psi}{\gamma + H_x}\right)^{\frac{1}{2}} . \quad \text{(A.17)}$$

The variances are respectively $\frac{1}{2}\psi^{-\frac{3}{2}}$ and $\frac{1}{2}\psi^{\frac{1}{2}}(\gamma + H_x)^{-\frac{3}{2}}$, so that the coefficient of variation is not constant with age as in the case of the gamma distribution, that is $\mathrm{cv}_x = 2^{-\frac{1}{2}}\ [\psi\,(\gamma + H_x)]^{-\frac{1}{4}}$. This indicates that the models based on these two distributions have fundamentally different properties. The inverse Gaussian (contrary to the gamma) would result in a population tending to become more homogeneous at older ages and then comprising mainly individuals with low frailty levels. Then, as Hougaard [1984] has pointed out "...the difference from ordinary life table methods is smaller for the inverse Gaussian than for the gamma distribution". He has also argued that, for a frailty model, this would be a more appropriate property, arising from the mortality selection process. However, this view was later challenged by Manton and Stallard [1984] (and subsequent work) who described the population aging process by a continuous-state continuous-time stochastic model, assuming that the effect of *diffusion* ("the random movement of persons to more extreme values of the physiological variables") would oppose the effect of selection, thus maintaining a constant coefficient of variation of the frailty variable.

Similarly to before, substituting the above equation (A.17) into the cohort force of mortality given by (2.10) yields the intrinsic relationship between the population hazard and the individual hazard function:

$$\overline{\mu}(x) = \mu_x \left(\frac{\psi}{\gamma + H_x}\right)^{\frac{1}{2}} . \quad \text{(A.18)}$$

In this inverse Gaussian case, we have considered only the Gompertz individual hazard rate. Transforming again the population hazard into a more suitable

format to facilitate the fitting process, we obtain:

$$\overline{\mu}(x) = \beta \cdot e^{p \cdot x} \left(\frac{\psi}{\gamma + \frac{\beta}{p}(e^{p \cdot x} - 1)} \right)^{\frac{1}{2}} = e^{p \cdot x} \left(\frac{\psi \cdot \beta^2 \cdot p}{(\gamma \cdot p - \beta) + \beta \cdot e^{p \cdot x}} \right)^{\frac{1}{2}}$$

$$= e^{p \cdot x} \left(\frac{\frac{\psi \cdot \beta^2 \cdot p}{\gamma \cdot p - \beta}}{1 + \frac{\beta}{\gamma \cdot p - \beta} e^{p \cdot x}} \right)^{\frac{1}{2}} = \frac{e^{-d + p \cdot x'}}{\sqrt{1 + e^{-b + p \cdot x'}}}, \tag{A.19}$$

where b, d and p are estimated directly from the fitting of the model and the β, ψ and γ can be computed as follows. As before we fix the mean frailty of the initial birth cohort to be 1, so that $\gamma = \psi$:

$$\hat{\beta} = \frac{e^{-\hat{d} - 40\hat{p}}}{\sqrt{1 + e^{-\hat{b} - 40\hat{p}}}}, \tag{A.20}$$

$$\hat{\gamma} = \hat{\psi} = \frac{e^{-\hat{d} - 40\hat{p}}}{e^{-\hat{b} - 40\hat{p}}} \frac{\sqrt{1 + e^{-\hat{b} - 40\hat{p}}}}{\hat{p}}, \tag{A.21}$$

and observe that we have $\beta = \overline{\mu}(0) = \mu_0$ as expected when $\overline{z}_0 = 1$.

We note that many authors have deduced the above formulae for the mean frailty for survivors (eqs. (A.7) and (A.17)) and for the population hazard functions (eqs. (A.8) and (A.18)) by applying the original method proposed by Vaupel et al [1979]. This approach could be useful for other marginal distributions of $\boldsymbol{Z}_x$. Thus, we can use the equation for the joint *p.d.f.* of age and frailty (2.5) to express the individual force of mortality at age x conditional on frailty level z in another form:

$$\mu(x \mid z) = \frac{f(x,z)}{S(x,z)}, \tag{A.22}$$

77Y which leads to a corresponding result for the unconditional cohort force of mortality considered for the whole range of z:

$$\overline{\mu}(x) = \frac{\overline{f}(x)}{\overline{S}(x)} = \frac{\int_0^\infty f(x,z)\,\mathrm{d}z}{\int_0^\infty S(x,z)\,\mathrm{d}z}$$

$$= \frac{\mu_x \cdot \int_0^\infty z \cdot e^{-z \cdot H_x} \cdot f_0(z)\,\mathrm{d}z}{\int_0^\infty e^{-z \cdot H_x} \cdot f_0(z)\,\mathrm{d}z}, \tag{A.23}$$

as in the development of (2.9).

Allowing, as before, for an appropriate *p.d.f.* of frailty among the cohort members at birth we can evaluate, in certain cases, the analytical relation between the population and individual hazard function at any current age x to which the members of the population might survive.

The transformed equations (A.10) and (A.19) (for the cases of gamma and inverse Gaussian frailty distributions, respectively) have the advantage of ensuring that

the multiplicands in the denominator are positive. For example, Congdon [1993] in fitting the Heligman – Pollard model which, for old ages, reduces to

$$q_x = \frac{G \cdot H^x}{1 + \kappa \cdot G \cdot H^x}, \tag{A.24}$$

obtains parameter estimates for $\hat{\kappa}$ which are negative (for females, Greater London, 1980 – 82). This would imply that $q_x \to \infty$ as $x \to \overline{x}$, where $1 + \kappa \cdot G \cdot H^{\overline{x}} = 0$.

c) **Other Distributions**

Aalen [1988] proposes a generalization of the models put forward by Hougaard [1984, 1986]. These lead to expressions for the population hazard rate of the following form

$$\overline{\mu}(x) = \frac{\mu_x}{(1 + \delta \cdot \psi^{-1} \cdot H_x)^{\psi}} \qquad \text{for } \psi,\ \delta \geq 0 \tag{A.25}$$

where the special case $\psi = 1$ corresponds to the gamma model, $\psi = \frac{1}{2}$ corresponds to the inverse Gaussian model, $\psi < 1$ corresponds to the stable distributions proposed by Hougaard [1986], $\psi \to 0$ corresponds to the degenerate case $\overline{\mu}(x) \to \mu_x$. The class of compound Poisson distributions would correspond to $\psi > 1$. Then, $\boldsymbol{Z}$ is either equal to 0 with positive probability or it is continuously distributed for $z > 0$. This model would not be appropriate for total mortality (i.e. there is no non–susceptibility to death!) but could be useful for particular diseases or causes of death — as shown by Aalen [1988] in the case of leukaemia.

d) **Other Choices for the Force of Mortality**

As noted in the text (section 2.2), some authors have proposed using the Weibull function for μ_x, for example Manton et al [1986]. This has not been pursued here because of the weight of evidence from the actuarial literature in favour of the Gompertz – Makeham family of mortality models for insurance – based populations.

e) **Other Structures**

Alternatives to the fundamental structure (equation (2.1)) suggested by Vaupel et al [1979] have been briefly considered. For example, we could mention:

$$\mu(x \,|\, Z = z) = f(z) \cdot \mu(x \,|\, 1) = f(z) \cdot \mu_x \tag{A.26}$$

where $f(\cdot)$ is a low order polynomial, like the linear form:

$$f(z) = f_0 + f_1 z\,. \tag{A.27}$$

Then the Gompertz – gamma model would lead to

$$\overline{\mu}(x) = f_0 \cdot \beta \cdot e^{p \cdot x} + \frac{f_1 \cdot \theta \cdot p \cdot \beta \cdot e^{p \cdot x}}{(\theta \cdot p - \beta) + \beta \cdot e^{p \cdot x}} \tag{A.9'}$$

to replace (A.9), and the Gompertz – inverse Gaussian model would lead to

$$\overline{\mu}(x) = f_0 \cdot \beta \cdot e^{p \cdot x} + f_1 \cdot \beta \cdot e^{p \cdot x} \left(\frac{\theta}{\theta + \frac{\beta}{p}(e^{p \cdot x} - 1)} \right)^{\frac{1}{2}} \qquad \text{(A.19')}$$

to replace (A.19). Preliminary exploratory analysis indicated that neither (A.9') nor (A.19') provides a good fit to the data sets investigated. Higher order polynomial choices for $f(z)$ have not been considered.

Jones [1998] explores the representation

$$\mu(x \mid Z = z) = z^{\gamma} \mu_x \qquad -\infty < \gamma < \infty \qquad \text{(A.28)}$$

for the transition intensity from the healthy state to the withdrawn state. The parameter γ is used to measure the impact of frailty on the "force of lapsation", with $\gamma = 0$ corresponding to independence. If $\gamma > 0$, then individuals with higher frailty would be more likely to lapse than those with lower frailty (and vice versa for $\gamma < 0$).

References

[1988] Aalen, O.O., "Heterogeneity in survival analysis", *Statistics in Medicine*, **vol. 7**, pp. 1121–1137.

[1959] Beard, R.E., "Appendix: Note on some mathematical mortality models", *CIBA Foundation Colloquia on Aging*, **vol. 5**, pp. 302–311.

[1971] Beard, R.E., "Some aspects of theories of mortality, cause of deaths analysis, forecasting and stochastic processes", *Biological aspects of Demography*, pp. 57–68, Edited by W. Brass, London, Taylor and Francis.

[1993] Benjamin, B. and Pollard, J.H., "The analysis of mortality and other actuarial statistics", *The Institute of Actuaries and the Faculty of Actuaries.*

[1955] Blumen, I., Kogan, M. and McCarthy, P.J., "The industrial mobility of labour as a probability process", *Cornell Studies of Industrial and Labour Relations*, **vol. vi**, Cornell University Press, Ithaca, New York.

[1992] Carriere, J.F., "Parametric models for Life Tables", *Transactions of the Society of Actuaries*, **vol. 44**, pp. 77–99.

[1998] Chang, S-C., "Using parametric statistical models to estimate mortality structure: The case of Taiwan", *Journal of Actuarial Practice*, **vol. 6**, pp. 197–219.

[1993] Congdon, P., "Statistical graduation in local demographic analysis and projection", *Journal of the Royal Statistical Society*, Series A, **vol. 156**, pp. 237–270.

[1994] Congdon, P., "Analyzing mortality in London: Life Tables with frailty", *Journal of the Royal Statistical Society*, Series D, **vol. 43**, pp. 277–308.

[1983] Cox, D.R., "Some remarks on overdispersion", *Biometrika*, **vol. 70**, pp. 269–274.

[1935] Cramer, H., and Wold, H., "Mortality variations in Sweden: a study in graduation and forecasting", *Skandinavisk Aktuarieidskrift*, **vol. 18**, pp. 161–241.

[1998] Damaskos, E., "Investigating models of frailty of mortality", *MSc Dissertation, Department of Actuarial Science & Statistics, City University*, London.

[1988] Forfar, D.O., McCutcheon, J.J. and Wilkie, A.D., "On graduation by mathematical formula", *Journal of the Institute of Actuaries*, **vol. 115**, pp. 1–149.

[1988] Forfar, D.O. and Smith, D.M., "The changing shape of English Life Tables", *Transactions of the Faculty of Actuaries*, **vol. 40**, pp. 98–134.

[1924] Gini, C., "Premières recherches sur la fécondabilité de la femme", *Proceedings of the International Mathematical Congress*, **vol. 2**, pp. 889–892.

[1872] Gompertz, B., "On one uniform law of mortality from birth to extreme old age, and on the law of sickness", *Journal of the Institute of Actuaries*, **vol. 16**, pp. 329–344.

[1996] Haberman, S. and Renshaw, A.E., "Generalized linear models and actuarial science", *Journal of the Royal Statistical Society*, Series D, **vol. 45**, pp. 407–436.

[1986] Harvey, A.C. and Durbin, J., "The effects of seat belt legislation on British road casualties: A case study in structural time series modelling", *Journal of the Royal Statistical Society*, Series A, **vol. 149**, pp. 187–227.

[1980] Heligman, L. and Pollard, J.H., "The age pattern of mortality", *Journal of the Institute of Actuaries*, **vol. 107**, pp. 49–80.

[1990] Hoem, J.M., "Identifiability in hazard model with unobserved heterogeneity: The compatibility of two apparently contradicting results", *Theoretical Population Biology*, **vol. 37**, pp. 124–128.

[1990] Horiuchi , S. and Coale, A.J., "Age patterns of mortality for older women: an analysis using the age–specific rate of mortality change with age", *Mathematical Population Studies*, **vol. 2**, pp. 245–267.

[1984] Hougaard, P., "Life Table methods for heterogeneous populations: Distributions describing heterogeneity", *Biometrika*, **vol. 71**, pp. 75–83.

[1986] Hougaard, P., "Survival models for heterogeneous populations derived from stable distributions", *Biometrika*, **vol. 73**, pp. 387–396.

[1998] Jones, B.L., "A model for analyzing the impact of selective lapsation on mortality" *North American Actuarial Journal*, **vol. 2, no. 1**, pp. 79–86.

[1979] Keyfitz, N. and Littman, G., "Mortality in heterogeneous population", *Population Studies*, **vol. 33**, pp. 333–342.

[1982] Keyfitz, N., "Choice of functions for mortality analysis: Effective forecasting depends on a minimum parameter representation", *Theoretical Population Biology*, **vol. 21**, pp. 329–352.

[1985] Keyfitz, N., "Applied mathematical demography", Second Edition, *Springer*, New York.

[1990] Lancaster, T., "The econometric analysis of transition data", *Cambridge University Press*, Cambridge.

[1976] Le Bras, H., "Lois de mortalité et age limité", *Population*, **vol. 31**, pp. 655–692.

[1959] Levinson, L.H., "A theory of mortality classes", *Transactions of the Society of Actuaries*, **vol. 11**, pp. 46–87.

[1867] Makeham, W. M., "On the law of mortality", *Journal of the Institute of Actuaries*, **vol. 13**, pp. 325–358.

[1980] Manton, K.G. and Stallard, E., "A stochastic compartment model representation of chronic disease dependence: Techniques for evaluationg parameters of partially unobserved age inhomogeneous stochastic processes", *Theoretical Population Biology*, **vol. 18**, pp. 57–75.

[1984] Manton, K.G. and Stallard, E., "Recent trends in mortality analysis", *Academic press, Inc.*, pp. 236–299.

[1981] Manton, K.G., Stallard, E. and Vaupel, J.W., "Methods for comparing the mortality experience of heterogeneous populations", *Demography*, **vol. 18**, pp. 389–410.

[1986] Manton, K.G., Stallard, E. and Vaupel, J.W., "Alternative models for the heterogeneity of mortality risks among the aged", *Journal of the American Statistical Association*, **vol. 81**, pp. 635–644.

[1989] McCullagh, P. and Nelder, J.A., "Generalized linear models", Second edition, *Chapman and Hall*, London.

[1989] McNown, R. and Rogers, A., "Time-series analysis forecasts of a parameterized mortality schedule", In "Advances in regional demography: Information, forecasts, models", Edited *P Congdon and P Batey*, pp. 107–123, *Bellhaven Press*, London.

[1997] Olshansky, S.J. and Carnes, B.A., "Ever since Gompertz", *Demography*, **vol. 34**, pp. 1–15.

[1932] Perks, W. F., "On some experiments in the graduation of mortality statistics", *Journal of the Institute of Actuaries*, **vol. 63**, pp. 12–40.

[1983] Ratkowsky, D.A., "Nonlinear regression modelling: A unified practical approach", *Marcel Dekker*, New York.

[1990] Ratkowsky, D.A., "Handbook of nonlinear regression models", *Marcel Dekker*, New York.

[1969] Redington, F.M., "An exploration into patterns of mortality", *Journal of the Institute of Actuaries*, **vol. 95**, pp. 243–298.

[1991] Renshaw, A.E., "Actuarial graduation practice and generalized linear and non-linear models" *Journal of the Institute of Actuaries*, **vol. 118**, pp. 295–312.

[1992] Renshaw, A.E., "Joint modelling for actuarial graduation and duplicate policies", *Journal of the Institute of Actuaries*, **vol. 119**, pp. 69–85.

[1995] Renshaw, A.E., "Graduation and generalized linear models: An overview", *Department of Actuarial Science & Statistics, City University*, London, Actuarial Research Paper No. 73.

[1997] Renshaw, A.E. and Haberman, S., "Dual modelling and select mortality", *Insurance: Mathematics and Economics*, **vol. 19**, pp. 105–126.

[2000] Sithole, T., Haberman, S. and Verrall, R.J., "An investigation into parametric models for mortality projections, with applications to immediate annuitants and life office pensioners", *Insurance: Mathematics and Economics*, **vol. 27**, pp. 285–312.

[1999] Thatcher, A.R., "The long–term pattern of adult mortality and the highest attained age", *Journal of Royal Statistical Society*, Series A, **vol. 162**, pp. 5–30.

[1988] Vaupel, J.W., "Inherited frailty and longevity", *Demography*, **vol. 25**, pp. 277–287.

[1999] Vaupel, J.W., Contribution to the discussion of "The long–term pattern of adult mortality and highest attained age", by A.R. Thatcher, *Journal of the Royal Statistical Society*, Series A, **vol. 162**, pp. 31–32.

[1979] Vaupel, J.W., Manton, K.G. and Stallard, E., "The impact of heterogeneity in individual frailty on the dynamics of mortality", *Demography*, **vol. 16**, pp. 439–454.

[1985a] Vaupel, J.W. and Yashin, A.I., "Heterogeneity's ruses: Some surprising effects of selection on population dynamics", *The American Statistician*, **vol. 39**, pp. 176–185.

[1985b] Vaupel, J.W. and Yashin, A.I., "The deviant dynamics of death in heterogeneous Populations", *Sociological Methodology*, pp. 179–211, Edited *N B Tuma, Jossey–Bas*, San Francisco.

[1988] Vaupel, J.W., Yashin, A.I. and Manton, K.G., "Debilitation's aftermath: Stochastic process models of mortality", *Mathematical Population Studies*, **vol. 1**, pp. 21–48.

[1999] Venables, W. N. and Ripley, B.D., "Modern applied statistics with S–Plus", *Springer–Verlag*, New York, Second Edition.

[1998] Wang, S.S. and Brown, R.L., "A frailty model for projection of human mortality improvement", *Journal of Actuarial Practice*, **vol. 6**, pp. 221–241.

[1977] Woodbury, M.A. and Manton, K.G., "A random–walk model of human mortality and aging", *Theoretical Population Biology*, **vol. 11**, pp. 37–48.

[1985] Yashin, A.I., Manton, K.G., and Vaupel, J.W., "Mortality and aging in a heterogeneous population: A Stochastic process model with observed and unobserved variables", *Theoretical Population Biology*, **vol. 27**, pp. 154–175.

[1994] Yashin, A.I., Vaupel, J.W. and Iachine, I.A., "A duality in aging: The equivalence of mortality models based on radically different concept", *Mechanisms of Aging and Development*, **vol. 74**, pp. 1–14.

[1995] Yashin, A.I., Vaupel, J.W. and Iachine, I.A., "Correlated individual frailty: An advantageous approach to survival analysis of bivariate data", *Mathematical Population Studies*, **vol. 5**, pp. 145–159.

[1997] Yashin, A.I. and Iachine, I.A., "How frailty models can be used for evaluating longevity limits: Taking advantage of an interdisciplinary approach", *Demography*, **vol. 34**, pp. 31–48.

CITY City University London

DEPARTMENT OF ACTUARIAL SCIENCE AND STATISTICS

Actuarial Research Papers since 2000

---

122. Booth P.M. and Cooper D.R. The Tax Treatment of Pensions. April 2000. 36 pages. ISBN 1 901615 42 1

123. Walsh D.E.P. and Rickayzen B.D. A Model for Projecting the number of People who will require Long-Term Care in the Future. Part I: Data Considerations. July 2000. 37 pages. ISBN 1 901615 43 X

124. Rickayzen B.D. and Walsh D.E.P. A Model for Projecting the number of People who will require Long-Term Care in the Future. Part II: The Multiple State Model. July 2000. 27 pages. ISBN 1 901615 44 8

125. Walsh D.E.P. and Rickayzen B.D. A Model for Projecting the number of People who will require Long-Term Care in the Future. Part III: The Projected Numbers and The Funnel of Doubt. July 2000. 61 pages. ISBN 1 901615 45 6

126. Cooper D.R. Security for the Members of Defined Benefit Pension Schemes. July 2000. 23 pages. ISBN 1 901615 45 4

127. Renshaw A.E. and Haberman S. Modelling for mortality reduction factors. July 2000. 32 pages. ISBN 1 901615 47 2

128. Ballotta L. and Kyprianou A.E A note on the □-quantile option. September 2000. ISBN 1 901615 49 9

129. Spreeuw J. Convex order and multistate life insurance contracts. December 2000. ISBN 1 901615 50 2

130. Spreeuw J. The Probationary Period as a Screening Device. December 2000. ISBN 1 901615 51 0

131. Owadally M.I. and Haberman S. Asset Valuation and the Dynamics of Pension Funding with Random Investment Returns. December 2000. ISBN 1 901615 52 9

132. Owadally M.I. and Haberman S. Asset Valuation and Amortization of Asset Gains and Losses in Defined Benefit Pension Plans. December 2000. ISBN 1 901615 53 7

133. Owadally M.I. and Haberman S. Efficient Amortization of Actuarial Gains/Losses and Optimal Funding in Pension Plans. December 2000. ISBN 1 901615 54 5

134. Ballotta L. □-quantile Option in a Jump-Diffusion Economy. December 2000. ISBN 1 901615 55 3

135. Renshaw A. E. and Haberman S. On the Forecasting of Mortality Reduction Factors. February 2001. ISBN 1 901615 56 1

136. Haberman S., Butt Z. & Rickayzen B. D. Multiple State Models, Simulation and Insurer Insolvency. February 2001. 27 pages. ISBN 1 901615 57 X

137. Khorasanee M.Z. A Cash-Flow Approach to Pension Funding. September 2001. 34 pages. ISBN 1 901615 58 8

138. England P.D. Addendum to "Analytic and Bootstrap Estimates of Prediction Errors in Claims Reserving". November 2001. 17 pages. ISBN 1 901615 59 6

139. Verrall R.J. A Bayesian Generalised Linear Model for the Bornhuetter-Ferguson Method of Claims Reserving. November 2001. 10 pages. ISBN 1 901615 62 6

140. Renshaw A.E. and Haberman. S. Lee-Carter Mortality Forecasting, a Parallel GLM Approach, England and Wales Mortality Projections. January 2002. 38 pages. ISBN 1 901615 63 4

141. Ballotta L. and Haberman S. Valuation of Guaranteed Annuity Conversion Options. January 2002. 25 pages. ISBN 1 901615 64 2

142 Butt Z. and Haberman S. Application of Frailty-Based Mortality Models to Insurance Data. April 2002. 65 pages. ISBN 1 901615 65 0

Statistical Research Papers

1. Sebastiani P. Some Results on the Derivatives of Matrix Functions. December 1995. 17 Pages. ISBN 1 874 770 83 2

2. Dawid A.P. and Sebastiani P. Coherent Criteria for Optimal Experimental Design. March 1996. 35 Pages. ISBN 1 874 770 86 7

3. Sebastiani P. and Wynn H.P. Maximum Entropy Sampling and Optimal Bayesian Experimental Design. March 1996. 22 Pages. ISBN 1 874 770 87 5

4. Sebastiani P. and Settimi R. A Note on D-optimal Designs for a Logistic Regression Model. May 1996. 12 Pages. ISBN 1 874 770 92 1

5. Sebastiani P. and Settimi R. First-order Optimal Designs for Non Linear Models. August 1996. 28 Pages. ISBN 1 874 770 95 6

6. Newby M. A Business Process Approach to Maintenance: Measurement, Decision and Control. September 1996. 12 Pages. ISBN 1 874 770 96 4

7. Newby M. Moments and Generating Functions for the Absorption Distribution and its Negative Binomial Analogue. September 1996. 16 Pages. ISBN 1 874 770 97 2

8. Cowell R.G. Mixture Reduction via Predictive Scores. November 1996. 17 Pages. ISBN 1 874 770 98 0

9. Sebastiani P. and Ramoni M. Robust Parameter Learning in Bayesian Networks with Missing Data. March 1997. 9 Pages. ISBN 1 901615 00 6

10. Newby M.J. and Coolen F.P.A. Guidelines for Corrective Replacement Based on Low Stochastic Structure Assumptions. March 1997. 9 Pages. ISBN 1 901615 01 4.

11. Newby M.J. Approximations for the Absorption Distribution and its Negative Binomial Analogue. March 1997. 6 Pages. ISBN 1 901615 02 2

12. Ramoni M. and Sebastiani P. The Use of Exogenous Knowledge to Learn Bayesian Networks from Incomplete Databases. June 1997. 11 Pages. ISBN 1 901615 10 3

13. Ramoni M. and Sebastiani P. Learning Bayesian Networks from Incomplete Databases. June 1997. 14 Pages. ISBN 1 901615 11 1

14. Sebastiani P. and Wynn H.P. Risk Based Optimal Designs. June 1997. 10 Pages. ISBN 1 901615 13 8

15. Cowell R. Sampling without Replacement in Junction Trees. June 1997. 10 Pages. ISBN 1 901615 14 6

16. Dagg R.A. and Newby M.J. Optimal Overhaul Intervals with Imperfect Inspection and Repair. July 1997. 11 Pages. ISBN 1 901615 15 4

17. Sebastiani P. and Wynn H.P. Bayesian Experimental Design and Shannon Information. October 1997. 11 Pages. ISBN 1 901615 17 0

18. Wolstenholme L.C. A Characterisation of Phase Type Distributions. November 1997. 11 Pages. ISBN 1 901615 18 9

19. Wolstenholme L.C. A Comparison of Models for Probability of Detection (POD) Curves. December 1997. 23 Pages. ISBN 1 901615 21 9

20. Cowell R.G. Parameter Learning from Incomplete Data Using Maximum Entropy I: Principles. February 1999. 19 Pages. ISBN 1 901615 37 5

21. Cowell R.G. Parameter Learning from Incomplete Data Using Maximum Entropy II: Application to Bayesian Networks. November 1999. 12 Pages ISBN 1 901615 40 5

22. Cowell R.G. FINEX : Forensic Identification by Network Expert Systems. March 2001. 10 pages. ISBN 1 901615 60X

23. Cowell R.G. When Learning Bayesian Networks from Data, using Conditional Independence Tests is Equivalant to a Scoring Metric. March 2001. 11 pages. ISBN 1 901615 61 8

Department of Actuarial Science and Statistics

Actuarial Research Club

The support of the corporate members

CGNU Assurance
Computer Sciences Corporation
English Matthews Brockman
Government Actuary's Department
HCM Consultants (UK) Ltd
KPMG
PricewaterhouseCoopers
Swiss Reinsurance
Watson Wyatt Partners

is gratefully acknowledged.

ISBN 1 901615 65 0